2021 DSI Annual Conference Proceedings

2021 DSI Proceedings Coordinator: Xun Xu (California State University, Stanislaus, USA)

ISBN: 978-0-578-62648-2
ISSN: 2471-884X

Table of Contents

DECISION SCIENCES INSTITUTE

Managing Concurrency in Projects under Stochastic Environments: Understanding Vaccine Development for Pandemics

Nagesh N. Murthy
Lundquist College of Business, University of Oregon
Email: nmurthy@uoregon.edu

Haikady N. Nagaraja
College of Public Health, The Ohio State University
Email: nagaraja.1@osu.edu

Hossein Rikhtehgar Berenji
College of Business, Pacific University
Email: hossein@pacificu.edu

ABSTRACT

Aggressive overlapping of activities associated with vaccine development, approval, and gearing up for its manufacturing and distribution during the COVID-19 pandemic has been one of the major levers to make effective vaccines available to the public, at "pandemic" or "warp" speed. We provide a framework to manage overlapping of stochastic tasks.

KEYWORDS: Project Management, Stochastic Tasks, Concurrency, COVID-19 Pandemic, Vaccine Development

INTRODUCTION

The pharmaceutical industry and governments worldwide are faced with an unprecedented uncertainty and urgency in developing vaccines as part of their disaster management strategy to deal with the COVID-19 pandemic. Aggressive overlapping of activities associated with vaccine development, approval, and manufacturing stages is being considered as one of the major levers to make effective vaccines available to the public at the earliest. A great majority of previous research has provided insights and guidance for managing concurrency (i.e., overlapping of activities) in projects to minimize project completion time by considering task duration to be deterministic (e.g., Ha and Porteus 1995, Roemer et al. 2000). Our research is motivated by the need to offer additional insights for managing concurrency via judicious overlapping to minimize project completion time in stochastic task environments. In particular, our motivation for addressing overlapping in stochastic task environments is well-illustrated by the need to understand and offer insights for managing the new and ambitious paradigm shift for the development of therapeutics and vaccines at "pandemic" or "warp" speed. We show that the interplay between stochastic overlap duration and the ensuing impact on expected project completion time manifests in ways that would not be apparent if one were to analyze optimal overlapping of activities by considering deterministic task duration. Our results indicate that stochastic overlapping in scenarios (illustrated in the context of vaccine development for COVID-19) that may be deemed ineffective to minimize project completion time when analyzed using deterministic task duration, can actually be beneficial when evaluated using stochastic task duration. We also examine implications of the extent to which resource flexibility in projects may permit tasks to be rescheduled during project planning, on the degree of concurrency and performance of project planning policies in a stochastic task environment. We develop a framework to gain insights by analyzing basic building blocks for planning projects. These building blocks comprised of 1-to-1, 1-to-n, andn-to-1 stochastic tasks can guide the development of heuristic for optimizing expected completion time in a stochastic project network. The implications for this research extend to managing projects in stochastic task environments in sectors well beyond the vaccine and therapeutics development industry. The emergence of disruptive core product or process technologies across a variety of industries is creating increased uncertainty in task duration and posing a major challenge for managing development of new to the world products and services in a speedy and timely manner.

REFERENCES

Ha AY, Porteus EL (1995) Optimal timing of reviews in concurrent design for manufacturability. Management Science41(9):1431–1447.

Roemer TA, Ahmadi R, Wang RH (2000) Time-cost trade-offs in overlapped product development. Operations Research48(6):858–865.

DECISION SCIENCES INSTITUTE

Driving or Deterring Pre-Recall Consumer Harm? An Examination of Institutional Investor Orientation

Catherine A. Helmuth
Central Michigan University
Email: helmu1c@cmich.edu

Brian L. Connelly
Auburn University
E-mail: bconnelly@auburn.edu

David J. Ketchen
Auburn University
E-mail: ketchda@auburn.edu

Christopher W. Craighead
University of Tennessee
E-mail: craighead@utk.edu

Kang Bok Lee
Auburn University
E-mail: kbl0009@auburn.edu

Sebastien Vendette
Central Michigan University
Email: vende1s@cmich.edu

ABSTRACT

Product recalls are a dangerous supply-chain disruption where defective goods often produce consumer harm before they are formally recalled. Product recall research, to date, has focused on supply-chain rather than corporate governance antecedents. To address this disparity, we theorize the firm's institutional investor orientation and deployed efforts to constraint shareholder power affects the level of consumer harm. In doing so, our results reveal dedicated and transient institutional investors deter and drive, respectively, pre-recall consumer harm. More profoundly, we find constitutional constraints on shareholder power increase pre-recall consumer harm suggesting such constraints enhance—not limit—the power of transient investors.

KEYWORDS: Corporate governance, Product recalls, Disruptions

DECISION SCIENCES INSTITUTE
A Decision Model For Circular Economy

Divya Choudhary
Indian Institute of Management Sambalpur, Odisha, India
divyachoudhary2626@gmail.com

"ABSTRACT"

Increasing number of organizations are incorporating circular economy (CE) practices in their business to sustain in the competitive market. CE is multifaceted and associated with various risks. The study aims to identify and evaluate the risks associated with CE practices. The literature addressing the risk aspects in the context of CE is very limited. Accordingly, a decision model is proposed to quantify and prioritize the identified CE risks. The decision model will facilitate the managers in the identification of critical areas that require immediate attention for the successful implementation of CE.

KEYWORDS: Circular Economy (CE), Decision Model, Risk Quantification, Risk Evaluation

INTRODUCTION

The business communities characterized by increasing global competition and complexities are facing rapidly intensifying risky environments recently (Heckmann et al. 2015). In context of supply chains, the increased outsourcing of activities in order to become more agile and lean is resulting in enhanced exposure to risk events (Tuncel & Alpan, 2010). However, many products move beyond the conventional supply chain perspective due to the value retained in them even at the end of their life (Prahinski & Kochabgolu, 2006). The set of processes, which incorporates the flow of products beyond conventional supply chain horizon for value reclamation, constitutes the Circular Economy (CE). The concept of recovery channels has gained recognition especially, in the last decade as organizations begin progressing towards sustainable practices owing to government pressure, consumer awareness, green image, competitiveness etc. (Krike et al, 2013). However, the scope of CE is greater than just addressing the legal and environmental responsibilities, as it offers great potential to cost reduction, revenue generation, customer retention and value addition etc. (Ullah & Sarkar, 2020). Furthermore, managing product returns at a societal level in an effective and cost-efficient way can help develop sustainable economies.

Despite the advantages, management of CE is quite complicated due to the involvement of multifunctional contextual environment of many return processes (Choudhary et al. 2021). Moreover, the flow of returns has very less visibility compared to the conventional supply chains, which results in surfacing of large number of uncertainties in the recovery channels (Bai & Sarkis, 2013). These ambiguities can trigger unplanned events resulting in unwanted consequences and thereby, making the recovery system vulnerable to risks. Therefore, in order to ensure efficient functioning of CE and to attain high economic as well as environmental performance, risk exposure of recovery operations needs to be optimized and managed (Singh & Aggarwal, 2018). In order to accomplish the same, a company is required to build expertise and knowledge, which needs to be codified, institutionalized and reinforced throughout all the recovery operations.

In addition, although considerable amount of research has been done in the domain of product recovery, the area of risk management is still immature and hence, can be considered as a research gap in CE dimension (Choudhary et al. 2021). Additionally, Govindan et al. (2015) stated in their review paper that risks and uncertainties have not been efficiently considered in most of the reverse logistics system and suggested it as a promising future research direction. In light of this, the research seeks to address the gap in literature regarding CE risk issues. Accordingly, the aim of study is to map and effectively understand the uncertainties existing in recovery channel followed by the quantitative analysis of risks. Furthermore, most of the risk analysis techniques applied in literature e.g. analytical hierarchy process, TOPSIS, failure mode and effect analysis etc. are insufficient to account for large amount of ambiguities and subjectivities (Liu et al., 2011). Hence, in order to assess the vague and complex structure of CE, the paper proposes an integrated risk analysis methodology by combining the fuzzy set theory (FST) and evidential reasoning algorithm (ER). The contribution of the paper is threefold: 1) Identification and discussion of risks present in CE; 2) Quantification and assessment of recovery channel risks in order to determine the high priority critical risks; 3) Introduction of an integrated risk analysis methodology by combining the Fuzzy Set Theory (FST) and Evidential Reasoning algorithm (ER).

LITERATURE REVIEW

CE has obtained recognition both as a research field and as a practice especially in the last decade (Rogers & Tibben-Lembke, 1999). The growing scarcity of resources, depleting landfill facilities, increasing rate of returns and legislative regulations enacted by the government are obligating the organizations to implement CE practices (Barker & Zabinsky, 2011). In addition, efficient management of returns provides a competitive edge to companies and offers an opportunity to combine environmental stewardship with monetary benefits through reduced resource investments (Krikke et al. 2013). CE can be characterized as an instrument that encompasses a set of recovery operations aimed at reclaiming the residual value of end-of-life, end-of-use, obsolete and warranty returns flowing backward from consumers to manufacturers and in turn enables the social and economical developments (Ullah & Sarkar, 2020). However, the flow of returned products, which represents a sizeable asset for many organizations, looses much of its strategic value in the recovery process due to the presence of various risks and uncertainties (Blackburn et al. 2004). The return channel for most of the products is relatively undeveloped, ambiguous and complex, which makes it susceptible to large number of disruptions (Farhani et al. 2019). Hence, in accordance with prevailing discussion, it is essential to explore the risks present in CE and mitigate them for improving the effectiveness of recovery practices.

In the light of prevailing discussion, for implementing the risk management practices in recovery channel, it is necessary to identify, understand and map the associated threats for enhancing the visibility within the system (Juttner et al. 2003). Thereafter, it is required to analyze the situations responsible for surfacing of the potential disruptions. It involves evaluation and quantification of the considered risk issues, which can be done based on their occurrence probability and impact (Tuncel & Alpan, 2010). In line with this, the study attempts to identify and prioritize the risks present in CE. Accordingly, structuring upon the literature of CE and supply chain risk management, some of the risks present in the various phases of CE have been identified as presented in Table 1.

After the initial recognition of risks in return channel, it is required to determine the most critical threats so that organizations can properly assign their resources for controlling such risks. Thus, fuzzy set theory and ERA are being used to analyze the acknowledged risks as described in the next section.

Table 1: Risks in Circular Economy

Risk in CE	Description
Collection risk (R1)	Risk associated with the location of collection points i.e. weather they are accessible to all range of customers willing to return products
Product Returns Forecast Risk (R2)	Risk due the inability to predict the time and amount of return products.
Network Design Risk (R3)	Disruptions caused due to inefficient design of reverse logistics routes and due to the un-optimized positioning of warehouses/ reprocessing centers
Marginal Value of Time Risk (R4)	Risk due to the inability to process the returned products within the desired time which can in turn lead to deterioration of residual value of returns.
Decision Making Risk (R5)	Risk associated with the selection of appropriate processing technique such as recycling, reuse, cannibalization, disposal etc. for recapturing the value of return products.
Material Recovery Risk (R6)	Risk of lower-than-expected recovery from the return products which can affect the plant's effective capacity.
Reprocessed Quality Risk (R7)	Quality related issues of remanufactured, repaired, secondary products and products manufactured from recycled raw material.
Recovery System Acceptance Risk (R8)	Risk related to consequences of not complying with the appropriate regulations (local and global) and uncertainty about changing return policies.
Market Risk (R9)	Uncertainty associated with the market response related to the resale of returned and reprocessed products.
Lack of Awareness Risk (R10)	Risk due to the deficient knowledge among the society regarding the concept of RSC and its significance.
Sustainable Regulations and Compliance Risk (R11)	Risk related to consequences of not complying with the appropriate regulations (local and global) and uncertainty about changing return policies.

METHODOLOGY

In order to have the complete awareness of threats across CE, it is further required to assess the risks for evaluating their priority (Shankar et al. 2018). Assessment of risks is quite intricate and difficult due to the involvement of large amount of preconceptions. Hence, the study proposes flexible and robust techniques such as fuzzy set theory (FST) and ER approach to quantify and analyze CE risks (Choudhary et al. 2020). Fuzzy set theory is very useful in modeling the linguistic assessments especially in case of vague and incomplete information (Wadhwa et al. 2009) whereas; ER approach has the superiority to handle the uncertainties and subjectivity in its inference procedure while avoiding the loss of substantial information (Mokhtari et al. 2012). The steps of the methodology to analyze the CE risks are as follows:

1. Identification of risks present in CE.
2. Linguistic assessment of risks.
3. Risk evaluation using Fuzzy Set Theory (FST).
4. Application of Belief Degrees for assessing risk priority levels in CE.
5. Assessment aggregation using Evidential Reasoning Algorithm (ERA).
6. Determination of Equivalent Risk Priority Score via Expected Utility Theory.

Linguistic Assessment of Risks

Risk issues are mostly evaluated across two dimensions including probability of occurrence and impact (Tuncel & Alpan, 2010). Furthermore, the impact of risk events will be examined across three aspects namely, operational impact (OI), environment & safety impact (E&SI) and financial impact (FI) (Saidi et al. 2014). Subsequently, the severity of risk events can be determined by calculating the risk priority level (L) based on the above mentioned dimensions using following equation (John et al. 2014).

$$L = P \times I \qquad (1)$$

Where, P is the probability that a risk will occur, I is the impact of the risk event on the concerned system and x represents the multiplication relationship among the two dimensions. This definition also implies that L will be a fuzzy number if P and I are depicted as fuzzy numbers. Accordingly, risk priority level of each risk issue associated with CE will be determined using Equation 1.

Application of FST for Risk Evaluation

The exercise to analyze risks often deals in both linguistic and numeric data and hence, it is essential to implement a methodology for the conversion of linguistic data into its corresponding numeric equivalent (Liu et al. 2011). Accordingly, fuzzy theory concepts can be utilized to quantify the expert opinions, which have been expressed qualitatively for assessing the risks in previous section. Further, a fuzzy membership function based on a five – scale method, adopted and tailored from Wadhwa et al.(2009), is employed in order to transform the linguistic assessment of risk parameters into TFNs with uniform distribution as shown in Table 2.

Table 2: Transformation of Linguistic Variables to Fuzzy Triangular Membership Functions

Scale	Probability (P)	Operational Impact (OI)	Environment & Safety Impact (E&SI)	Financial Impact (FI)	Membership Function
1	Very Low (VL)	Very Low (VL)	Very Low (VL)	Very Low (VL)	(0.00,0.00,0.25)

2	Low (L)	Low (L)	Low (L)	Low (L)	(0.00,0.25,0.50)
3	Medium (M)	Medium (M)	Medium (M)	Medium (M)	(0.25,0.50,0.75)
4	High (H)	High (H)	High (H)	High (H)	(0.50,0.75,0.10)
5	Very High (VH)	Very High (VH)	Very High (VH)	Very High (VH)	(0.75,100,1.00)

Additionally, as per Equation 1, $TFN_{(PI)}$ can be calculated as shown below in Equation 2-3:

$$TFN_{(PI)} = TFN_{(P)} \otimes TFN_{(I)} = (x_P * x_I,\ y_P * y_I,\ z_P * z_I) \quad (2)$$

Where,

$$TFN_{(I)} = Avg.(TFN_{(OI)} + TFN_{(E\&SI)} + TFN_{(FI)}) = \left(\frac{x_{OI}+x_{ESI}+x_{FI}}{3}, \frac{y_{OI}+y_{ESI}+y_{FI}}{3}, \frac{z_{OI}+z_{ESI}+z_{FI}}{3} \right) \quad (3)$$

Moreover, based on Equations 2 and 3, FST can be utilized to categorize the risks as explained in Table 3. In Table 3, risk levels have been defined on the basis of triangular fuzzy numbers and the centroid value, K has been used for analyzing the risk issues. The centroid value, K can be determined as follows:

$$K = \frac{1}{3}(x + y + z) \quad (4)$$

Table 3: Description of Qualitative Scale used for Risk Analysis based on TFNs:

Qualitative Scale for risk analysis	Description of evaluation variables		Membership Function	Centroid Values (K)
	Probability	Impact		
Very Low (VL): Acceptable risk	VL (0.00,0.00,0.25)	VL (0.00,0.00,0.25)	(0.0,0.0,0.0625)	0.020
Low (L): Tolerable risk but should be reduced if feasible.	L (0.00,0.25,0.50)	L (0.00,0.25,0.50)	(0.00,0.0625,0.25)	0.104
Medium (M): Significant risk that should be reduced considering cost involved.	M (0.25,0.50,0.75)	M (0.25,0.50,0.75)	(0.0625,0.25,0.5625)	0.292
High (H): Severe risk that must be reduced.	H (0.50,0.75,1.00)	H (0.50,0.75,0.10)	(0.25,0.5625,1.00)	0.604
Very High (VH): Highly severe risk, which must be reduced and controlled immediately.	VH (0.75,100,1.00)	VH (0.75,100,1.00)	(0.5625,1.00,1.00)	0.854

Further, the computed $TFN_{(PI)}$ for each risk issue needs to be transformed into fuzzy risk that can be represented in an integrated space of discourse, which can be supplied as an input to ER approach for evaluating CE risks. Hence, fuzzy risk and the associated belief degrees will be computed as discussed in the next section.

Application of Belief Degrees for Assessing Risk Priority Levels

A belief degree is the measure of potency to which a judgment is believed to be true and it can be exactly 100% or less. After the transformation of linguistic assessment of risks into TFNs, it is required to convert the fuzzy rating of risk parameters into belief degree structures having the similar set of evaluation grades (Mokhtari et al. 2012). Accordingly, the product of fuzzy probability ($TFN_{(P)}$) and fuzzy impact ($TFN_{(I)}$) i.e. $TFN_{(PI)}$ needs to be transformed into fuzzy risk assigned with belief degrees having the same set of linguistic variables (John et al., 2014). The following set of evaluation grades will be used to assess the risk priority level for each individual risk issue:

$$R = \{R_1, R_2, R_3, R_4, R_5\}$$
$$= \{\text{Very Low, Low, Medium, High, Very High}\}$$

Further, the transformation of $TFN_{(PI)}$ into fuzzy risk, R defined by 5 linguistic terms constituting a normalized fuzzy set i.e. belief structure, can be done as illustrated in Figure 1 and Table 4.

Table 4: Example of Transformation of $TFN_{(PI)}$ into Belief Degree

$TFN_{(PI)}$	0.29,0.63,1.00				
Grade	VL	L	M	H	VH
R_L	0.00	0.36	0.78	0.80	0.40
R	0.00	0.15	0.33	0.34	0.17

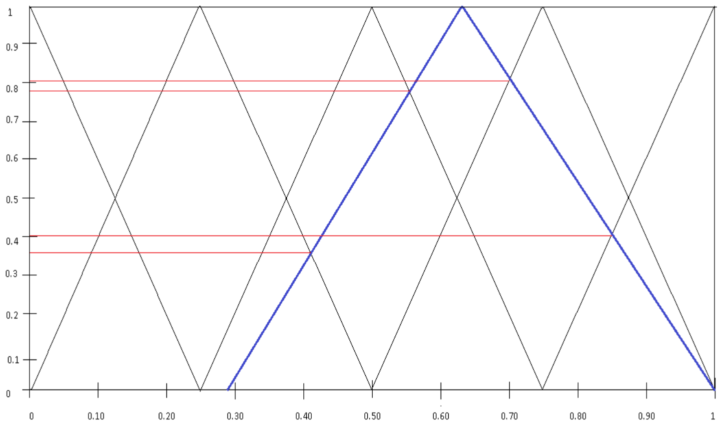


Figure 1: Example of Transformation of $TFN_{(PI)}$ into 5 Non-Normalized Grades

As mentioned before, the belief structure determined above will now be utilized as an input to ER approach, which will aggregate the belief degrees in order to determine the risk priority level for each individual risk as explained in the next section.

Application of Evidential Reasoning Algorithm for Assessment Aggregation

The ER approach is developed based on the Dempster–Shafer theory also known as D-S theory of evidence, which was initially generated by Dempster and later developed by Shafer (Shafer, 1976). However, ER algorithm has been applied for risk analysis in various domains, in this paper it will be utilized to assess the CE risks by following the explained pathway (Yang, 2001).

If R represents a class of five linguistic risk expressions that has been obtained by the aggregation of two subclasses R_1 and R_2 provided by two different experts, then R, R_1 and R_2 can be individually represented by:

R = (α^1 “Very Low”, α^2 “Low”, α^3 “Medium”, α^4 “High”, α^5 “Very High”)
R_1= (α_1^1 “Very Low”, α_1^2 “Low”, α_1^3 “Medium”, α_1^4 “High”, α_1^5 “Very High”)
R_2= (α_2^1 “Very Low”, α_2^2 “Low”, α_2^3 “Medium”, α_2^4 “High”, α_2^5 “Very High”)

Where, linguistic expressions are allied with their corresponding belief degrees. Assume w_1 and w_2 ($w_1 + w_2 = 1$) as the normalized relative weights of the two experts involved in the assessment of product recovery risks. Further, let M_i^n (n= 1,2,3,4 and 5; i= 1or 2) represents the degree to which i[th] subset (R_i) support the hypothesis that risk priority level has been assessed to five risk expressions. Then M_1^n and M_2^n can be calculated as:

$$M_1^n = w_1 \times \alpha_1^n; \qquad M_2^n = w_2 \times \alpha_2^n \tag{5}$$

Suppose S_1 and S_2 are the lingering belief degree values unassigned for M_1^n and M_2^n (n= 1,2,3,4 and 5). Then, S_1 and S_2 can be obtained as follows (Liu et al. 2011):

$$S_1 = \bar{S}_1 + \tilde{S}_1; \qquad S_2 = \bar{S}_2 + \tilde{S}_2 \tag{6}$$

Where, $\bar{S}_m$ (m= 1 or 2) corresponds to the degree to which the other expert can have a role in the evaluation whereas, $\tilde{S}_m$ (m = 1 or 2) represents the probable incompleteness in the subsets R_1 and R_2. They can be expressed as given below:

$$\bar{S}_1 = 1- w_1 = w_2 ; \qquad \bar{S}_2 = 1- w_2 = w_1 \tag{7}$$
$$\tilde{S}_1 = w_1\left(1-\sum_{n=1}^{5}\alpha_1^n\right), \qquad \tilde{S}_2 = w_2\left(1-\sum_{n=1}^{5}\alpha_2^n\right) \tag{8}$$

Let $\alpha^{n'}$(n= 1,2,3,4 and 5) corresponds to the non-normalized degree that the evaluation of risks is established to each of the five linguistic risk expressions due to the aggregation of assessments provided by the expert 1 and 2. Further, suppose that the non-normalized unassigned belief degree left after the assignment of belief to the five risk expressions as a result of the aggregation of assessments provided by the expert 1 and 2 is represented by $S_U^{'}$. The recursive ER algorithm can be depicted as follows (Yang et al. 2009):

$$\alpha^{n'} = K\left(M_1^n M_2^n + M_1^n S_2 + M_2^n S_1\right)$$
$$\bar{S}_U^{'} = K\left(\bar{S}_1\bar{S}_2\right) \tag{9}$$
$$\tilde{S}_U^{'} = K\left(\tilde{S}_1\tilde{S}_2 + \tilde{S}_1 S_2 + S_1\tilde{S}_2\right)$$
$$K = \left[1-\sum_{j=1}^{5}\sum_{\substack{l=1\\ l\neq j}}^{5} M_1^j M_2^l\right]^{-1} \tag{10}$$

Further, the combined degree of belief, α^n generated by the aggregation of assessments provided by the experts for each risk issue and the unassigned belief degree, α_U representing the overall incompleteness in evaluation are calculated as follows (Mokhtari et al., 2012):

$$\alpha^n = \alpha^{n'} / (1-\bar{S}_U^{'}) \text{ (n= 1,2,3,4 or 5)} \tag{11}$$
$$\alpha_U = \tilde{S}_U^{'} / (1-\bar{S}_U^{'})$$

The aforementioned process corresponds to the combination of two subsets only. If more subsets are required to be synthesized, then the outcome of aggregation of any two subsets can further be combined with the other subset in accordance with the steps of above algorithm. Similarly, the judgments of various experts can also be combined using this algorithm as illustrated later in the paper.

Determination of Equivalent Risk Priority Score via Expected Utility Theory

In this section, the distributed assessment of risk priority level determined for each disruption through ER algorithm will be transformed to an equivalent numerical value using expected utility theory for the ease of comparison and evaluation of risks. The single value obtained after the transformation can be termed as risk priority score (RPS) corresponding to the priority level of each individual risk issue. The utilities of evaluation terms are equidistantly distributed in the normalized utility space and can be determined as follows (John et al., 2014):

$$u(Z_n) = (H_n - 1)/(H_N - 1) \quad (n= 1,2\text{-------}N) \tag{12}$$

H_n is the rating of the evaluation term under consideration (Z_n) and H_N is the rating of the most preferred linguistic evaluation grade (Z_N). Further, the aggregated distribution assessment of risk issues in terms of risk priority level obtained by combining the expert judgments can be represented as:

$$S(L_i) = \{ (Z_n, \alpha_n(L_i)), n = 1,2,3,\ldots\ldots, N\} \tag{13}$$

Where, α_n represents the belief degree that the risk priority level L_i is assessed to evaluation grade Z_n and α_u denotes the remaining unassigned degree of belief. However, if all the evaluations are complete i.e. α_u =0, then the minimum, average and maximum utilities of $S(L_i)$ will be same. Hence $u(S(L_i))$ can be determined as follows (Yang et al., 2009):

$$u(S(L_i)) = \sum_{n=1}^{N} \alpha_n\, u(\mathrm{Zn}) \tag{14}$$

The $u(S(L_i))$ calculated above will give us the RPS for that particular risk and is only utilized for characterizing an evaluation and not for aggregation.

Application of Methodology for Analyzing Risks

This section will illustrate the application of the described methodology for assessing risk factors present in the reverse channel operations. The judgments of three experts with experience of about 12-15 years in the concerned field have been utilized for the evaluation process as well as for the identification of risks along with literature review. The identified risk issues are shown in Table 1. After the identification, risk factors are evaluated by the three experts based on probability of occurrence (P) and impact (I), which include operational impact (OI), environment & safety Impact (E&SI) and financial impact (FI) using the linguistic terms shown in Table 2. The assessment is presented in Table 5.

Table 5: Linguistic Assessment of CE risks

Risks Issues	Expert 1				Expert 2				Expert 3			
	P	OI	E&SI	FI	P	OI	E&SI	FI	P	OI	E&SI	FI
R1	VH	H	H	H	M	L	M	H	H	M	VH	H
R2	H	H	M	VH	H	H	L	VH	M	H	L	H
R3	H	H	M	L	H	H	M	H	H	H	H	H
R4	VH	H	H	H	VH	M	H	VH	H	L	L	H
R5	H	H	M	VH	H	VH	M	VH	M	L	H	H
R6	VH	H	H	H	H	M	M	H	H	M	M	M
R7	H	H	M	L	H	H	L	VH	M	M	M	M
R8	H	H	M	L	M	H	M	H	M	M	M	L
R9	VH	H	H	H	H	VH	L	VH	H	H	M	H
R10	H	H	M	L	H	VL	VL	VH	H	H	H	H
R11	VH	H	H	H	VH	H	M	VH	H	H	H	H

Further, the linguistic assessment of risk parameters shown in Table 5 is transformed into equivalent TFNs based on Table 2. Fuzzy impact $(TFN)_I$ is calculated as described in Equation 3. Further, risk priority level for each risk is obtained in the form of $(TFN)_{PI}$ in accordance with Equation 2 as shown in Table 6.

Table 6: Risk Priority Level in the Form of $(TFN)_{PI}$ & Intersection Results of Risk Issues

Risk Issues	Expert 1						Expert 2						Expert 3					
	$(TFN)_{PI}$	R_L					$(TFN)_{PI}$	R_L					$(TFN)_{PI}$	R_L				
		VL	L	M	H	VH		VL	L	M	H	VH		VL	L	M	H	VH
R1	(0.38,0.75,1.00)	0.00	0.20	0.60	1.00	0.00	(0.06,0.25,0.56)	0.43	0.00	0.56	0.11	0.00	(0.29,0.56,0.92)	0.00	0.41	0.89	0.69	0.28
R2	(0.25,0.56,0.92)	0.00	0.42	0.88	0.68	0.28	(0.21,0.50,0.83)	0.08	0.54	0.00	0.57	0.14	(0.08,0.29,0.63)	0.37	0.92	0.65	0.22	0.00
R3	(0.13,0.38,0.75)	0.22	0.76	0.81	0.41	0.00	(0.21,0.50,0.92)	0.08	0.54	0.50	0.63	0.26	(0.25,0.56,1.00)	0.00	0.45	0.89	0.70	0.37
R4	(0.38,0.75,1.00)	0.00	0.20	0.60	1.00	0.00	(0.17,0.44,0.75)	0.00	0.20	0.60	0.00	0.40	(0.08,0.31,0.67)	0.36	0.87	0.69	0.28	0.00
R5	(0.25,0.56,0.92)	0.00	0.42	0.88	0.68	0.28	(0.29,0.63,0.91)	0.00	0.36	0.80	0.77	0.31	(0.08,0.29,0.63)	0.37	0.92	0.65	0.22	0.00
R6	(0.38,0.75,1.00)	0.00	0.20	0.60	1.00	0.00	(0.17,0.44,0.83)	0.16	0.64	0.90	0.52	0.12	(0.13,0.38,0.75)	0.25	0.74	0.81	0.41	0.00
R7	(0.13,0.38,0.75)	0.22	0.76	0.81	0.41	0.00	(0.21,0.50,0.83)	0.08	0.54	0.00	0.57	0.14	(0.06,0.25,0.56)	0.43	0.00	0.56	0.11	0.00
R8	(0.13,0.38,0.75)	0.22	0.76	0.81	0.41	0.00	(0.10,0.33,0.69)	0.32	0.84	0.72	0.32	0.00	(0.04,0.21,0.44)	0.50	0.92	0.40	0.00	0.00
R9	(0.38,0.75,1.00)	0.00	0.20	0.60	1.00	0.00	(0.25,0.56,0.83)	0.00	0.45	0.89	0.64	0.16	(0.21,0.50,0.92)	0.08	0.54	0.00	0.63	0.26
R10	(0.13,0.38,0.75)	0.22	0.76	0.81	0.41	0.00	(0.13,0.25,0.50)	0.33	0.00	0.50	0.00	0.00	(0.25,0.56,1.00)	0.00	0.45	0.89	0.70	0.37
R11	(0.38,0.75,1.00)	0.00	0.20	0.60	1.00	0.00	(0.31,0.58,0.75)	0.00	0.20	0.60	0.00	0.40	(0.25,0.56,1.00)	0.00	0.45	0.89	0.70	0.37

Next, step involves transforming the fuzzy rating of risk priority level into corresponding belief degree structure as explained in Figure 1 and Table 4. The intersection results for the risk issues are shown in Table 6, which are then further normalized. Finally, the risk assessments provided by three experts are synthesized for each individual recovery system risk issue by applying the evidential reasoning algorithm as explained in section 3.4 with the help of Equations 5-11. Since, all the three experts have profound knowledge of the concerned area and approximately same year of experience, they have been assigned equal weights in the aggregation calculations. The results of aggregation obtained by ERA are shown in Table 7. However, it is more realistic and easier to evaluate the risk factors present in a system based on a single value. Therefore, utility value of each linguistic grade and the overall utility associated with the distributed assessment of risk is calculated to obtain a crisp value of RPS for each risk issue using Equations 12-14 and are presented in Table 6. Further, based on Table 2 and RPS, risks are categorized as shown in Table 7.

Table 7: Categorization of Risks Based on RPSs

Risk Issues	Aggregated Assessment via ERA (α_n)					RPS	Rank	Categorization
	VL	L	M	H	VH			
R1	0.175	0.060	0.477	0.263	0.024	0.475	8	High
R2	0.064	0.379	0.163	0.319	0.075	0.490	7	High
R3	0.038	0.268	0.325	0.276	0.092	0.529	5	High
R4	0.033	0.206	0.456	0.154	0.151	0.545	3	High
R5	0.034	0.231	0.377	0.269	0.089	0.537	4	High
R6	0.053	0.239	0.389	0.297	0.022	0.499	6	High
R7	0.146	0.291	0.222	0.292	0.049	0.452	9	High
R8	0.151	0.435	0.311	0.104	0.000	0.342	11	Medium
R9	0.010	0.212	0.275	0.435	0.069	0.585	2	High
R10	0.205	0.117	0.543	0.105	0.030	0.410	10	Medium
R11	0.000	0.148	0.466	0.196	0.190	0.607	1	High

The computed values of priority scores and the equivalent classes determined for the risk issues provides ground for enhancing the performance of recovery operations by identifying the critical areas that require improvement and control.

CONCLUSIONS & DISCUSSIONS

Emerging sustainable practices and growing green concerns have encouraged the concept of CE in recent years. In addition, market pressure, government regulations and economical benefits are other reasons owing to which organizations are adopting recovery practices. However, successful accomplishment of CE goals is quite difficult since, a large number of risks are involved in recovery practices. Hence, as discussed earlier, for the efficient adoption of recovery operations, it is essential to map the associated risks in order to have the visibility across CE (Choudhary et al. 2021). For the same, FST and ER algorithm have been used to measure and categorize the risk events. Unlike conventional risk management techniques, this approach has the capability to model the precise data, vague judgments and uncertainties of a complex system in a unified manner, while preserving the original characteristics of multiple attributes.

Based on the results obtained in Table 7, none of the risk has been categorized as very low priority risk, which infers that all the considered risk issues can significantly impact the functioning of CE. It is worth mentioning that all these values have been obtained from the synthesis of expert opinions. Among all the considered risks, R11 occupies the first position and thus, holds the highest priority in CE. It suggests that organizations adopting the recovery practices are facing a major challenge in complying with the take back obligations and due to the changing legislative policies (Brito & Dekker, 2003, Atasu & Boyaci, 2010). Hence, management needs to emphasize more on restructuring their recovery practices in accordance with the government imposed regulations to curb the associated risk. Further, R9 possesses the second highest priority, which implies that organizations are unable to balance the demand with returns and it is further worsening due to the customer's reluctance in purchasing secondary products (Guide & Jayaram, 2000; Choudhary et al. 2020). Thus, managers must focus on enhancing the flexibility of recovery operations in order to balance the demand with the returns. Simultaneously, organizations should attempt to expand their customer base by making the secondary products more appealing through adopting practices such as technology upgrade, lowering the prices compared to new products and proper marketing etc.

Additionally, the suggested methodology and the obtained results of this study can be further extended and modified by the managers according to the specific characteristics of their organizations or products. However, as all the results are based on the expert's opinions working in the electronic industries, the generalization of the findings in other industries is uncertain. In future work, a risk index can be developed to measure the vulnerability of CE considering the interactions among the risks with the help of various techniques such as analytical hierarchy process, graph theory etc.

REFRENCES

Atasu, A., & Boyaci, T. (2010). Take-back legislation and its impact on closed-loop supply chains. *Wiley Encyclopedia of Operations Research and Management Science*. Available http://dx.doi.org/10.1002/9780470400531.eorms0869.

Barker, T. J., & Zabinsky, Z. B. (2011). A multicriteria decision making model for reverse logistics using analytical hierarchy process. *Omega* 39: 558-573.

Bai, C., & Sarkis, J. (2013). Flexibility in Reverse Logistics: a Framework and Evaluation Approach. *Journal of Cleaner Production* 47: 306-318.

Blackburn, Joseph, D., Guide, V., Gilvan, C. S., & Van Wassenhove, L. (2004). Reverse Supply Chains for Commercial Returns. *California Management Review* 46: 6-22.

Brito, M. P. and Dekker, P. (2003). A Framework for Reverse Logistics. *Erasmus Research Institute of Management Report Series*, Research in Management.

Choudhary, D., Shankar, R., & Choudhary, A. (2020). An integrated approach for modeling sustainability risks in freight transportation systems. *Risk Analysis, 40*(4), 858–883.

Choudhary, D., Choudhary, A., Shankar, R., & Hicks, C. (2021). Evaluating the risk exposure of sustainable freight transportation: a two-phase solution approach. *Annals of Operations Research*, 1-35.

Farahani, S., Otieno, W. & Barah, M., (2019). Environmentally friendly disposition decisions for end-of-life electrical and electronic products: The case of computer remanufacture. *Journal of Cleaner Production*, *224*, 25-39.

Govindan, K., Soleimani, H., & Kannan, D. (2015). "Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future. *European Journal of Operation Research* 240 (3): 603-626.

Guide, V.D.R., & Jayaraman, V. (2000). Product acquisition management: current industry practice and a proposed framework. *International Journal of Production Research* , 38: 3779–3800

Heckmann, I., Comes, T., & Nickel, S. (2015). A Critical Review on Supply Chain Risk–Definition, Measure and Modeling. *Omega* 52: 119-132.

John, A. et al. (2014). An integrated fuzzy risk assessment for seaport operations. *Safety Sciences* 68: 180-194.

Jüttner, U., Peck, H., & Christopher, M. (2003). Supply chain risk management: Outlining an agenda for future research. *International Journal of Logistics: Research and Applications* 6 (4): 197–210.

Krikke, H., Hofenk, D., & Wang, Y. (2013). Revealing An Invisible Giant: a Comprehensive Survey into Return Practices Within Original (Closed-Loop) Supply Chains. *Resources, Conservation and Recycling* 73: 239–250.

Liu, H. et al. (2011). Failure Mode and Effects Analysis using Fuzzy Evidential Reasoning Approach and Grey Theory. *Expert System with Application* 38: 4403-4415.

Mokhtari, K., Ren, J., Roberts, C., & Wang, J. (2012). Decision Support Framework for Risk Management on Sea Ports and Terminals using Fuzzy Set Theory and Evidential Reasoning Approach. *Expert System with Application* 39: 5087-5103.

Prahinski, C., & Kocabasoglu, C. (2006). Empirical Research Opportunities in Reverse Supply Chains. *Omega* 34: 519 – 532.

Rogers, D.S. & Tibben-Lembke, R.S. (2001). An examination of reverse logistics practices. *Journal of Business Logistics* 22: 129-148.

Saidi, E., Anvaripour, B., Jaderi, F., & Nabhani, N. (2014). Fuzzy risk modeling of process operations in the oil and gas refineries. *Journal of Loss Prevention in Process Industries* 30: 63-73

Shafer, G., (1976). A Mathematical Theory of Evidence. Vol. 1. Princeton University Press, New Jersey

Shankar, R., Choudhary, D., & Jharkharia, S. (2018). An integrated risk assessment model: A case of sustainable freight transportation systems. *Transportation Research Part D: Transport and Environment,* *63,* 662–676.

Singh, R.K. & Agrawal, S., (2018). Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach. *Management of Environmental Quality: An International Journal*, *29*(3), 427-443

Tuncel, G., & Alpan, G. (2010). "Risk Assessment and Management for Supply Chain Networks: A Case Study." *Computers in Industry* 61: 250–259.

Ullah, M. & Sarkar, B., (2020). Recovery-channel selection in a hybrid manufacturing-remanufacturing production model with RFID and product quality. *International Journal of Production Economics*, *219*, 360-374.

Wadhwa, S., Madaan, J., & Chan, F. T. S. (2009). Flexible Decision Modeling of Reverse Logistics System: a Value Adding MCDM Approach for Alternative Selection. *International Journal of Robotics and Computer-Integrated Manufacturing* 25(2): 460–469.

Yang, J. B. (2001). Rule and utility Based Evidential Reasoning Approach for Multi-Attribute Decision Analysis Under Uncertainties. *European Journal of Operational Research*, 131, 31–61.

Yang, Z., Wang, J., Bonsall, S., & Fang, Q., (2009). Use of Fuzzy Evidential Reasoning in Maritime Security Assessment. *Risk Analysis* 29 (1): 95–120.

DECISION SCIENCES INSTITUTE

How Sustainable Supply Chain Management Affects Environmental Performances: a Longitudinal Analysis of CDP Data

Alice Madonna
Department of Management, Information and Production Engineering, University of Bergamo
Email: alice.madonna@unibg.it

Albachiara Boffelli
Department of Management, Information and Production Engineering, University of Bergamo
Email: albachiara.boffelli@unibg.it

Matteo Kalchschmidt
Department of Management, Information and Production Engineering, University of Bergamo
Email: matteo.kalchschimdt@unibg.it

ABSTRACT

The study sets out to evaluate the effect of Sustainable Supply Chain Management (SSCM) on improving companies' environmental performance. A longitudinal empirical analysis was performed on panel data provided by the Carbon Disclosure Project (CDP) Supply Chain Program. The analyses showed that supply chain engagement, corporate governance and risk management, directly and interactively, affect multiple aspects of environmental performances. The most compelling conclusions revolved around the need to take a step further in managing for sustainability, especially concerning more effective collaboration with the engaged partners.

KEYWORDS: Socially Responsible Operations, Panel studies, Framework

INTRODUCTION

A decade ago, corporate sustainability and the disclosure thereof allowed those who practiced it to gain competitive advantage by increasing their collaborations with several partners and changing their business model to fit the novel sustainability practices and targets (Kiron, Kruschwitz, Haanaes, & Velken, 2012; Luo, Lan, & Tang, 2012). Nowadays, carbon disclosure is becoming mandatory to participate in the markets (Agarwal, Giraud-Carrier, & Li, 2018; Serafeim, 2020). An in-depth and reliable disclosure often results in a trustworthiness signal, thereby leading to a competitive advantage (Villena & Dhanorkar, 2020). Serafeim (2020) developed an empirical study examining more than 3000 firms during the first months of the 2020 pandemic financial crisis and observed that the companies perceived as more responsible had less negative stock returns than their competitors. This study further highlighted the need to embrace and embed the sustainability topic in the company's culture instead of barely considering it a list of boxes to tick through highly standardized ESG practices. On the same page, Benn, Edwards, and Williams (2018) emphasize the need to reshape 'business as usual' to secure better outcomes for society and the planet. Elkington (2018) affirms that action has been taken, especially from business leaders, which can supposedly influence lower tier supply chain partners (Gold, Seuring & Beske 2009); however, this does not always happen, either way, it is not enough.
According to the *Phase Model*, a framework that classifies organisations by their investment in the sustainability topic, the first steps towards sustainability are compliance with the regulations and the efficiency of the processes (Benn et al., 2018). The former entails reducing the risk to incur potential liabilities by adhering to environmental laws and signalling the interest in employees' engagement, while the latter seeks to reduce any avoidable cost by implementing value-adding strategies rather than cost-reduction (Benn et al., 2018). If this

framework was applied in a supply chain scenario, Green Supply Chain Management (GSCM) could achieve the first steps. GSCM seeks to reduce the possible negative impact of a supply chain by rethinking the product's whole life cycle (Agarwal et al., 2018). Overall, GSCM aims to enhance internal and external practices' efficiency by utilizing at maximum capacity each material resource, preventing waste, promoting recycling, and choosing the firm's partners more carefully. The implementation of GSCM is motivated by governments regulations, suppliers and market drivers (Zhu, Sarkis, & Lai, 2007; Walker, Di Sisto, & McBain, 2008; Tay, Abd Rahman, Aziz, Sidek, & Humanity, 2015).

Though all these supply chain management (SCM) practices aim to improve the firm's environmental performance, they principally address environmental issues, ignoring the economic and social aspects that could be entailed. Considerably broader is the scope of Sustainable Supply Chain Management (SSCM), whose purpose is to create, protect and increase long-term environmental, social and economic value. SSCM has a threefold set of targets, aligning with the triple bottom line (TBL) concept, a framework that addresses "profit, people and planet" related issues (Elkington, 1994). Saeed and Kersten (2019) summarised the literature regarding SSCM drivers to conclude that these practices are driven externally by societal, regulatory bodies and market pressures and internally by corporate strategy, organizational culture, resources and characteristics. The role of SSCM is to achieve both economic and non-economic targets and thereby create long-term value (Pagell & Wu, 2009; Tate, Ellram, & Kirchoff, 2010). Observing its role in the supply chain context, SSCM is comparable to the *Phase Model* strategic proactivity, which aims to reshape the firms' businesses, looking for competitive advantage and leadership. Acknowledging the critical role that SSCM has towards the development of more sustainable SCs, several authors in the extant literature have remarked the need for an integrated assessment for supply chains to evaluate them as whole entities, both internally and externally, instead of assessing the different components separately (Farooque, Zhang, Thürer, Qu, & Huisingh, 2019; Mahapatra, Schoenherr, & Jayaram, 2021; Tsai, Bui, Tseng, Ali, Lim, & Chiu, 2021). Therefore, this study sets out to contribute to this gap.

Finally, a firm becomes *Sustainable* when it achieves environmental best practices and engages its stakeholders and society to assist them in the process towards sustainability. Furthermore, stakeholders are engaged to strive for human sustainability (Benn et al., 2018). The path towards sustainability outlined in the *Phase Model* is the same that Serafeim (2020) illustrated through examples of virtuous business leaders, which again is the same as the one utilized by the Carbon Disclosure Project (CDP), the source of data for this study, to evaluate its participants (CDP, 2020b).

Carbon Disclosure Project (CDP) is one of the globally acknowledged corporations that provide sustainability ratings. Since 2012, investors and experts have recognized it as one of the best environmental ratings for its quality and usefulness (SustainAbility, 2012; 2020). CDP offers a series of questionnaires to companies, cities, investors, and governments to disclose their sustainability practices and results in a standardized format. Over the last two decades, CDP significantly raised the attention on environmental issues, creating an unparalleled level of engagement (CDP, 2020c). Today, CDP is regarded as the world's largest database on climate-change-related data (Damert, Feng, Zhu, & Baumgartner, 2018), gathering nearly $100 trillion of partners investor assets, nearly 10.000 reporting companies, and more than 800 reporting cities (CDP, 2020c). CDP's reporting initiatives proposed to companies address climate change and water and forest security (CDP, 2020a). These programs are issued with a questionnaire; in particular, the Climate Change questionnaire investigates the firms' business strategy, internal and external management and environmental impact. CDP ranks its respondents based on the survey response's depth and completeness, thus positioning the responding company in one of the stages towards environmental stewardship: disclosure, awareness, management and leadership (CDP, 2020b).

The Climate Change program can be accessed by firms through different channels: voluntarily, under buyer request (Supply Chain Program) or under investor request (Investor Program). Furthermore, the response to the questionnaire can either be public, thus published on the

CDP website, open for anyone to read, or private, thus being accessible only by buyers ad investors that requested the participation. Since the participation can be voluntary, the questionnaires' answers could be misleading and report only positive information. Nevertheless, several authors proved the reliability of the data disclosed on the respondents' heterogeneity, the disclosure repeatability, the trustworthiness of negative answers (Dobler, 2008; Cho, Lee, & Pfeiffer, 2013; Ott, Schiemann, & Günther, 2017) and the public scrutiny that public participation is subjected to (Mahapatra et al., 2021).

Notwithstanding the activities and recognitions of CDP, corporate sustainability disclosed through these questionnaires has not gained relevant consideration from Supply Chain Management (SCM) academics.

Since CDP participants are rated according to their environmental performance, this study sets out to investigate the impact of SSCM on enhancing the firm's environmental performance. Moreover, firms struggle to find a business case for sustainability even though it is well acknowledged that investing in sustainable innovations can improve the firms' financial performance, not only by attracting new investors or buyers but also generating cost savings from the enhanced efficiency achieved (Whelan & Fink, 2016). Seen this fact, the present study aims to evaluate the impact of SSCM on environmental performances and the sustainability business case: monetary savings generated by the investments in emission reduction activities. To do this, empirical analysis has been developed on the CDP data gathered through their *Supply Chain Program* (CDP-SCP).

Finally, several authors in the extant literature have remarked the need for an integrated assessment for supply chains to evaluate them as a whole entity, both internally and externally, instead of different assessments for its different components (Farooque et al., 2019; Mahapatra et al., 2021; Tsai et al., 2021); therefore, this study sets out to provide a contribution towards this gap.

In the remainder of this paper, the literature background we relied on to develop our hypotheses is first presented. Then, the methodology is explained, along with the sample and the measures. The model specification is then presented. The main findings and their discussion follow suit. Finally, the main conclusions, including limitations and future research, end the paper.

LITERATURE BACKGROUND AND HYPOTHESES DEVELOPMENT

Seuring and Müller (2008) defined SSCM as "management of material, information and capital flows as well as cooperation among companies along the supply chain while taking goals from all three dimensions of sustainable development" (Seuring & Müller, 2008, p 1700). Saeed and Kersten (2019) summarised and classified its drivers. The internal drivers are clustered in corporate strategy, organizational culture, resources and characteristics. Particular relevance is given to the top management commitment (*corporate strategy*), as it is necessary to raise the sustainability topic's urgency and to shift the firm's target towards the TBL (González-Benito & González-Benito, 2008). Concerning *organizational culture*, the companies need to closely monitor the business code of conduct as regulators highly control these practices to ensure they remain in line with their guidelines (Saeed & Kersten, 2019). The correct allocation of *organizational resources* aids the companies to pursue efficiently the initiatives engaged (Haverkamp, Bremmers, & Omta, 2010; Giunipero, Hooker, & Denslow, 2012; Schrettle, Hinz, Scherrer -Rathje, & Friedli, 2014). Finally, *organisational characteristics* are crucial to identify appropriate goals to reach and expectations to meet (Zhu & Sarkis, 2007; Haverkamp et al., 2010; Tate et al., 2010; Ayuso, Roca, & Colomé, 2013)

The external drivers to SSCM are clustered in regulatory, societal and market pressure. Coping with these external pressure is fundamental for companies to avoid possible liabilities or disadvantages (Schrettle et al., 2014). Particular relevance is given to suppliers and customers as their behavior in terms of carbon disclosure is capable of setting a non-written rule for the value chain participation. (Villena & Dhanorkar, 2020). Furthermore, buyers are crucial to establishing appropriate goals, while collaborating with suppliers is fundamental to achieve them (Tate et al., 2010; Gualandris & Kalchschmidt, 2014).

Prior research relying on the CDP questionnaires is relatively scarce, particularly in the SSCM field. Most authors considered the CDP surveys merely as a source of environmental performance and to check firms' compliance with market requests (Luo et al., 2012; Jira & Toffel, 2013; Luo, Tang, & Lan, 2013; Stanny, 2013; Gallego-Álvarez, García-Sánchez, & Da Silva Vieira, 2014; Ben-Amar & Mcilkenny, 2015; Blanco, Caro, & Corbett, 2016; Depoers, Jeanjean, & Jérôme, 2016; Guenther, Guenther, Schiemann, & Weber, 2016; Giannarakis, Zafeiriou, & Sariannidis, 2017; Ott et al., 2017; Li, Huang, Ren, Chen, & Ning, 2018; Luo, Tang, & Peng, 2018; Hsueh, 2019; Luo, 2019; Mateo-Márquez, González-González, & Zamora-Ramírez, 2019). These pieces of information are just partial if compared to the complexity and depth of the investigation the questionnaire pursues. Only a few authors dug in-depth into the questionnaires. Dahlmann, Branicki, and Brammer (2017) focused on organizational behavior, using information regarding environmental performance incentives to investigate how these are reflected in corporate and environmental performance. Fabrizio and Kim (2019) developed a linguistic study, investigating whether the degree of linguistic obfuscation in business strategy answers effectively secures a better rating by CDP. Villena and Dhanorkar (2020) and Damert et al. (2018) observed drivers that could moderate different pressures on a supplier. The first author used the CDP questionnaire to establish supplier transparency and the latter to determine the supplier's perception of risks and opportunities.
The remainder of this section delineates the major dimensions considered in this study, as well as the hypotheses formulated to investigate the drivers of SSCM by extensively relying on CDP data.

Stakeholder engagement
According to the stakeholder theory (Freeman, 1984), a firm's environmental practices could involve relevant suppliers and customers as stakeholders (Henriques & Sadorsky, 1999; Hsu, Choon Tan, Hanim Mohamad Zailani, & Jayaraman, 2013; Damert et al., 2018). Hahn, Reimsbach, and Schiemann (2015) summarise this theory in the context of environmental disclosure, claiming that providing information on carbon performance can be explained as a response to the pressure applied to the firm by its stakeholders, including suppliers.
Furthermore, building on institutional theory, which alleges that firms pursue profits and the fulfilment of institutional requirements (DiMaggio & Powell, 1983), Jira and Toffel (2013) claim that buyers can influence their suppliers' behavior by leading them towards carbon disclosure. Suppliers suffer the normative pressure generated by their buyers (industry leaders) and respond by complying with its expectations to meet their buyers' supplier selection criteria, namely carbon disclosure (DiMaggio & Powell, 1983; Villena & Dhanorkar, 2020). Given that focal firms cannot always influence the complete chain, first tiers suppliers are pressured to diffuse interest towards sustainability within their suppliers, particularly within those suppliers that focal firms are not in touch with, thus transferring the same pressures they are suffered to lower tiers suppliers (Wilhelm, Blome, Bhakoo, & Paulraj, 2016).
Therefore, it appears that the relevant stakeholders are able to stir firms towards carbon disclosure, whose quality is driven by the emission levels (Luo, 2019). Hence, the first research hypothesis is the following:

H1: The engagement of supply chain partners affects the firm's environmental performance.

Corporate governance
According to the agency theory, CSR should be developed exclusively to enhance the firm's profit; otherwise, it brings in contrast shareholders' and board's interest (Friedman, 1970). Therefore, corporate governance and carbon performance should be negatively associated. Nonetheless, with the rise of corporate sustainability (Luo et al., 2012), climate change has become a relevant topic for stakeholders. According to Chan, Watson, and Woodliff (2014), firms with more structured governance are more likely to behave more responsibly, being able to respond and address the needs of non-financially oriented stakeholders. More recently, Luo and Tang (2020) allege a positive relationship between corporate governance and carbon

performance; thus, the present study aims at evaluating this hypothesis utilizing data from a different source.
The feature of corporate governance taken into consideration in this study is the institution of incentives to reward the management of climate-change-related issues. Climate change incentives are a proxy to whether corporate sustainability is developed: providing them requires first risk analysis on the firm's operations, then identification and measurement capabilities to implement possible solutions and gauge their outcomes (Villena & Dhanorkar, 2020). The extant literature has proven that providing employees with a reward system effectively aligns their objectives with those of the organization (Matt & George, 1998; Tosi & Greckhamer, 2004; Bruce, Buck, & Main, 2005; Gerhart, Rynes, & Fulmer, 2009). Notably, Dahlmann et al. (2017) demonstrated that better-structured incentives lead to significant environmental performance improvement. Compared to the last-mentioned study, the present research aims to narrow the study on managerial incentives, assessing whether their presence is sufficient to cause environmental performance improvements. The hypothesis is defined as:

H2: The presence of climate-change managerial incentives positively affects the firm's environmental performance.

Risk management
McKinsey, along with International surveys observed by BSR, reports that climate-change-related issues could severely impact the firms' financial performance, even up to 70% of earnings (Bonini & Swartz, 2014). Seen the gauge of these damages, firms need to implement solid risk management strategies to manage the long-term risks that climate change poses (Whelan & Fink, 2016).
Prospect theory is a psychological theory based on the premise of risk aversion that explains decisions in uncertain conditions (Kahneman & Tversky, 1979; Kahneman, Tversky, MacLean, & Ziemba, 2000). In organizational studies, this theory reveals that issues identified as risks or opportunities receive different reactivity and treatment (Jackson & Dutton, 1988). The literature in the environmental context claims that a relation exists between an unbalanced identification of risks instead of opportunities and more unsatisfactory sustainability performances and ratings (Fiegenbaum & Thomas, 1988; Elijido-Ten, 2017). Elijido-Ten (2017) proved this relationship utilizing a sustainability rating influenced by the firm's carbon performance. Hence, the present study seeks to identify a relationship between risks and opportunities identification and environmental performance, defining the hypotheses as:
H3: The identification of climate-change-related risks negatively affects the firm's environmental performance.
H4: The identification of climate-change-related opportunities positively affects the firm's environmental performance.

METHODOLOGY

Sample
For this analysis, data have been observed and extracted from the CDP-SCP for the years between 2013 and 2018. Previous years' questionnaires were not included in the analysis as they did not include all the information utilized. The sample of respondents comprehends each listed company that has responded publicly to the questionnaire in any of the considered years, independently from its size, country or industry sector. These criteria led to 3,398 observations from a population of 1,178 different firms from 42 countries that adhered to CDP either once or perennially. Figure 1 and Figure 2 show respectively the number of respondents by total years of participation and year of participation. As shown in Figure 1, many of the respondents (more than 350) participated in the program for just one year.

Figure 1: Number of respondents by the number of years of disclosure

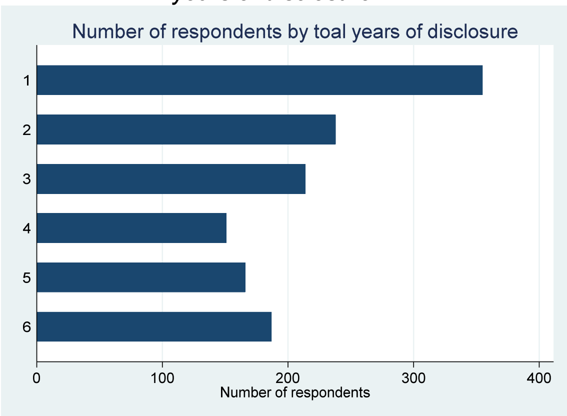
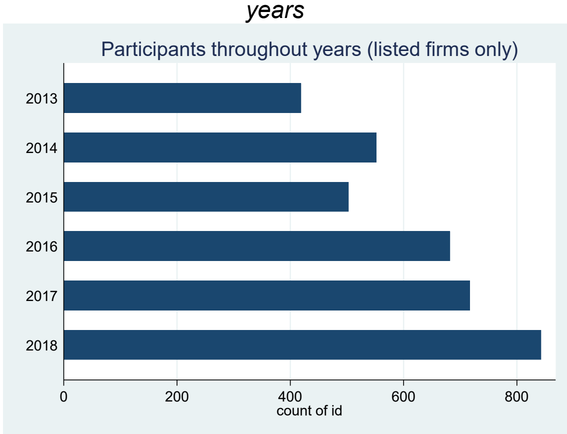


Figure 2: Number of respondents throughout the years



Measures

The present study has observed as dependent variables environmental performances in the form of CO_{2eq} emissions and monetary savings generated by emission reduction activities. Carbon performance has been measured through each firm's emission levels (Carbon emissions are negatively correlated to carbon performance, so an increase in the emission levels is reflected in worse carbon performance and vice-versa). The total emissions have been calculated by summing the Scope 1 and Scope 2 emission values (Dahlmann et al., 2017) reported through CDP. Scope 3 was not considered since it is not mandatory to report it in the CDP questionnaire and, therefore, can be unreliable (Huang, Weber, & Matthews, 2009; Blanco et al., 2016). Total savings have been calculated by summing the savings generated by each of the investments required to implement the reported initiatives. This monetary value was converted to US dollars to be compared by retrieving the mean exchange rates from FactSet. The performance values have been weighted by the total sales of each company, retrieved from ORBIS, to account for the scale factor (Gallego-Álvarez et al., 2014). Finally, both the variables have been log-transformed to reduce the skew.

The independent variables utilized accounted for the engagement of supply chain partners on climate change issues, the presence of climate change-related incentives and the number of climate change-related risks and opportunities identified. Given the premises on the CDP-SCP, the sample analyzed in this study is not randomized, but it is biased towards buyer-supplier collaboration. Indeed, several respondents specified that either supplier or buyer collaboration meant answering the questionnaire itself.

Two sets of dummy variables were created to identify whether the respondent was engaging with its suppliers and customers. The presence of climate change incentives was measured with a binary variable: the value of 1 accounted for the presence of the observed condition, a value of 0 for its absence, and a missing value for an unanswered question.

Risk management has been observed through the number of climate-change-related risks that, according to the respondent, could potentially have a significant financial or strategic impact on the firm. The value has been transformed with a square root transformation to reduce the skew.

The control variables observed are the firm's dimension and industry sector to account for exogenous phenomena (Jira & Toffel, 2013; Villena & Dhanorkar, 2020). The firm's dimension was measured through the number of employees and geographical extension, as the number of countries, the respondent indicated to generate emission. To account for the latter, a categorical variable has been created to identify four groups of firms whose branches are located exclusively in one country, in two to nine countries, in ten to forty-nine countries and in over fifty countries. The number of employees was retrieved from ORBIS and has been log-transformed to reduce the skew.

Descriptive statistics
Table 1 below summarizes how the variable names have been abbreviated and called in further sections of this study. Table 2 and Table 3 report descriptive statistics for the continuous and dummy variable utilized, while Table 4 reports descriptive statistics for the categorical variable of the industry sector.

Table 1: Variables' abbreviations index

Variable name	Abbreviation
Number of employees	EMPL
Geographical extension	NC
Industrial sector	NACE
Corporate governance: managerial incentives	CG_INC
Supply chain engagement: suppliers	ENG_SC_S
Supply chain engagement: customers	ENG_SC_C
Total identified risks	TR
Carbon Equivalent Emissions	EMS
Environmental savings from LCSCM practices	SAV

Table 2: Continuous variables descriptive statistics

Variable	N. Observations	Mean	SD	Min	Max
NC	3,398	1.86	1.127	1	4
EMPL	2,796	10.051	1.491	0	13.356
EMS	2,529	-9.728	1.652	-17.403	1.614
SAV	1,455	-11.104	4.615	-26.7	1.032
TR	2,596	2.259	1.204	0	8.944

Table 3: Dummy variables descriptive statistics

Variable	N. Observations	Mean	SD	Count 1	Count 0
CG_INC	3,398	.809	.388	2,959	692
ENG_SC_S	3,398	.653	.464	2,364	1,177
ENG_SC_C	3,398	.601	.478	2,184	1,357

Table 4: Industry sector control variable structure

	Overall		Between		Within
NACE	Freq.	Percent	Freq.	Percent	Percent
1	2,029	59.71	721	61.21	99.93
2	1,369	40.29	458	38.88	99.89
Total	3,398	100.00	1,179	100.08	99.92
			(n = 1,178)		

MODEL SPECIFICATIONS
The present study developed two series of panel-data models to test the formulated hypotheses, one for each of the dependent variables observed. The firms who responded to the questionnaire were the individuals *i*, and the year of participation was the time unit *t*. Panel models account for two error components: the idiosyncratic error term $\varepsilon_{i,t}$, and the so-called

individual heterogeneity α_i which captures time-invariant unobservable effects (Verbeek, 2017).
The complete models (equations (1) and (2)) evaluated the control variables, the independent variables' direct effect and their indirect effect through the moderation of other variables' effect simultaneously. Submodels analyzing exclusively two independent variables and their moderation have been observed. For each dependent variable, the significant independent variables and the moderations thereof have been finally grouped. The significant relationships have further been investigated for causality with panel-specific Granger causality Wald tests: none of the relationships was proven to be causal, as six time frames are at times not sufficient to determine causality (Greene, 2003).

$$\begin{aligned} EMS_{i,t} = {} & \varepsilon_{i,t} + \alpha_i + \beta_1 EMPL_{i,t} + \beta_2 NC_{i,t} + \beta_3 SIC\ M\ NM_{i,t} \\ & + \beta_4 CG\ INC_{i,t} + \beta_5 CG\ INC_{i,t-1} + \beta_6 ENG\ SC\ S_{i,t} + \beta_7 ENG\ SC\ S_{i,t-1} \\ & + \beta_8 ENG\ SC\ C_{i,t} + \beta_9 ENG\ SC\ C_{i,t-1} \\ & + \beta_{10} TR_{i,t} + \beta_{11} TR_{i,t-1} + \beta_{12}\ TO_{i,t} + \beta_{13} TO_{i,t-1} \\ & + \beta_{15} ENG\ SC\ S_{i,t} \cdot CG\ INC_{i,t} + \beta_{16} ENG\ SC\ S_{i,t-1} \cdot CG\ INC_{i,t-1} \\ & + \beta_{17} ENG\ SC\ C_{i,t} \cdot CG\ INC_{i,t} + \beta_{18} ENG\ SC\ C_{i,t-1} \cdot CG\ INC_{i,t-1} \\ & + \beta_{19} ENG\ SC\ S_{i,t} \cdot TR_{i,t} + \beta_{20} ENG\ SC\ S_{i,t-1} \cdot TR_{i,t-1} \\ & + \beta_{21} ENG\ SC\ C_{i,t} \cdot TR_{i,t} + \beta_{22} ENG\ SC\ C_{i,t-1} \cdot TR_{i,t-1} \\ & + \beta_{23} ENG\ SC\ S_{i,t} \cdot TO_{i,t} + \beta_{24} ENG\ SC\ S_{i,t-1} \cdot T0_{i,t-1} \\ & + \beta_{25} ENG\ SC\ C_{i,t} \cdot TO_{i,t} + \beta_{26} ENG\ SC\ C_{i,t-1} \cdot TO_{i,t-1} \end{aligned} \tag{1}$$

$$\begin{aligned} SAV = {} & \varepsilon_{i,t} + \alpha_i + \beta_1 EMPL_{i,t} + \beta_2 NC_{i,t} + \beta_3 SIC\ M\ NM_{i,t} \\ & + \beta_4 CG\ INC_{i,t} + \beta_5 CG\ INC_{i,t-1} + \beta_6 ENG\ SC\ S_{i,t} + \beta_7 ENG\ SC\ S_{i,t-1} \\ & + \beta_8 ENG\ SC\ C_{i,t} + \beta_9 ENG\ SC\ C_{i,t-1} \\ & + \beta_{10} TR_{i,t} + \beta_{11} TR_{i,t-1} + \beta_{12}\ TO_{i,t} + \beta_{13} TO_{i,t-1} \\ & + \beta_{15} ENG\ SC\ S_{i,t} \cdot CG\ INC_{i,t} + \beta_{16} ENG\ SC\ S_{i,t-1} \cdot CG\ INC_{i,t-1} \\ & + \beta_{17} ENG\ SC\ C_{i,t} \cdot CG\ INC_{i,t} + \beta_{18} ENG\ SC\ C_{i,t-1} \cdot CG\ INC_{i,t-1} \\ & + \beta_{19} ENG\ SC\ S_{i,t} \cdot TR_{i,t} + \beta_{20} ENG\ SC\ S_{i,t-1} \cdot TR_{i,t-1} \\ & + \beta_{21} ENG\ SC\ C_{i,t} \cdot TR_{i,t} + \beta_{22} ENG\ SC\ C_{i,t-1} \cdot TR_{i,t-1} \\ & + \beta_{23} ENG\ SC\ S_{i,t} \cdot TO_{i,t} + \beta_{24} ENG\ SC\ S_{i,t-1} \cdot T0_{i,t-1} \\ & + \beta_{25} ENG\ SC\ C_{i,t} \cdot TO_{i,t} + \beta_{26} ENG\ SC\ C_{i,t-1} \cdot TO_{i,t-1} \end{aligned} \tag{2}$$

Before running the analyses, the dataset was searched for the presence of first-order autocorrelation, heteroscedasticity and stationarity. Autocorrelation was sought through the Wooldridge test (Drukker, 2003), resulting in the acceptance of no first-order autocorrelation for both models. Heteroscedasticity was evaluated through the Breush-Pagan/Cook-Weisberg test (Breusch & Pagan, 1979; Cook & Weisberg, 1983), resulting in the constant variance hypothesis's rejection claiming heteroscedasticity for the emissions and rejection for the savings. To account for heteroscedasticity, the estimated variance-covariance matrix considered error term clustered by individual. Finally, the presence of stationarity was investigated: the Dickey-Fuller test (Dickey & Fuller, 1979) has been run on each of the dependent variables, resulting in the rejection of the hypotheses that the tested variable is stationary. Finally, correlations between the error term in each model and one or more independent variables have been accounted for utilizing the Hausman test (Hausman, 1978). This test accepted for both models the hypothesis strict exogeneity, no correlation between the error, therefore fixed effect estimators were used. This estimator exploits only the within-groups variability in the data; thus, the effect of a given change in $x_{i,t}$ on $y_{i,t}$ is the same regardless of whether the change is along the time or the cross-section dimension (Verbeek, 2017).

Another preliminary test evaluated the presence of multicollinearity, whose presence inflates the standard errors, thus jeopardizing the reliability of the estimation coefficients and the significance thereof. The main causes of multicollinearity could be two: the presence of a large number of dummy variables and the presence of variables at two time frames (t and t-1). To test the presence of multicollinearity between the independent variables, the value inflation factors (VIFs) were calculated: the mean VIFs in model (1) and (2) are 3.86 and 3.97 respectively, with worrying values for TO being nearly 8, but below the action threshold value of 10 (Belsley, 1980; Greene, 2003). These results revealed a moderate correlation between two or more independent variables, though not severe enough to require an intervention.

RESULTS

This section describes the results obtained: a first analysis observed the complete models represented in equations (1) and (2), whose results are reported in Table 5. The model on emissions shows that incentives have a negative direct effect on the dependent variable (CG_INC, p-value 0.026), and the presence of customer engagement weakens this effect (ENG_SC_C # CG_INC, p-value 0.018). Furthermore, the one-year lagged total number of risks positively affects the emission (L.TR, p-value 0.094). These results support H1, claiming an effect of supply chain partner engagement on the performance, H2, claiming a positive effect of incentives on the performance, and H3, claiming an association between risk identification and less favorable performance. This model does not show any support to H3.

The complete model on savings shows that the geographical extension of a firm positively affects the dependent variable; additionally, incentives have a positive direct effect on emissions (CG_INC, p-value 0.014), and the presence of customer engagement weakens this effect (ENG_SC_C # CG_INC, p-value 0.068). These results support H1, claiming an effect of supply chain partner engagement on the performance, and H2, claiming a positive effect of incentives on the performance. This model does not show any support to H3 or H3.

The second analyses level observed a pair of independent variables and the moderation thereof. Table 6 reports the results on emissions, while Table 7 on savings.

The results on the model (1.1) highlight a significant positive direct effect of supplier engagement (ENG_SC_S, p-value 0.047) on emission, which is weakened by the presence of incentives (ENG_SC_S # CG_INC, p-value 0.066). Model (1.2) shows that the collaboration with suppliers also has a lagged direct positive effect on emissions (L.ENG_SC_S, p-value 0.039), which again is weakened by the lagged total risks identified (L.ENG_SC_S # L.TR, p-value 0.024). Model (1.3) shows that collaboration with customers has a lagged direct negative effect on emissions (L.ENG_SC_C, p-value 0.057), which is weakened by the lagged total risks identified (L.ENG_SC_C # L.TR, p-value 0.056). These significant variables and moderations are grouped and tested simultaneously in the model (1.7): supplier collaboration loses its direct effect on the performance (ENG_SC_S), but it maintains the lagged effect, the presence of incentives gain significance (CG_INC, p-value 0.097) but loses the moderating effect on supplier engagement (ENG_SC_S # CG_INC). The lagged customer engagement and the moderation of total risks on both the supply chain engagement variables hold their significance.

These results support H1, claiming an effect of supply chain partner engagement on the performance, H2, claiming a positive effect of incentives on the performance, and H3, claiming an association between risk identification and less favorable performance. These models do not show any support to H3.

Table 5: Complete panel models analyses

	Emissions (1) [b]	Savings (2) [b]
EMPL	-0.337	-0.486
2.NC	0.090	**1.052***
3.NC	0.088	**1.281***
4.NC	0.091	**1.369***
2.SIC_M_NM	0.002	0.848
CG_INC	**-0.252****	**4.737****
L.CG_INC	0.008	0.197
ENG_SC_S	0.227	1.158
L.ENG_SC_S	0.282	2.026
ENG_SC_C	-0.014	3.250
L.ENG_SC_C	-0.350	-1.350
TR	-0.138	0.535
L.TR	**0.095***	0.682
TO	0.254	-0.017
L.TO	-0.127	-1.542
ENG_SC_S # CG_INC	0.196	0.177
L.ENG_SC_S # L.CG_INC	0.082	-4.431*
ENG_SC_C # CG_INC	**0.209****	**-3.847***
L.ENG_SC_C # L.CG_INC	0.047	-0.513
ENG_SC_S # TR	0.184	-0.725
L.ENG_SC_S # L.TR	-0.086	-1.029
ENG_SC_C # TR	0.034	0.687
L.ENG_SC_C # L.TR	0.013	0.692
ENG_SC_S # TO	-0.288	0.492
L.ENG_SC_S # L.TO	-0.072	2.800**
ENG_SC_C # TO	-0.124**	-0.582
L.ENG_SC_C # L.TO	0.153	-0.576
Number of observations	1092	568
R-Square	0.073	0.115

* p<0.10, ** p<0.05, *** p<0.01

The submodels analyzing savings show that the presence of incentives have a direct negative effect on the performance (model (2.1): CG_INC, p-value 0.042) and their lagged values weaken the positive effect that lagged supplier engagement has on the performance (model (2.4): L.ENG_SC_S, p-value 0.054; L.ENG_SC_S # L.CG_INC, p-value 0.091). Finally, model (2.5) shows that lagged customer engagement positively affects savings (L.ENG_SC_C, p-value 0.030) though being weakened by the moderation of total opportunities (L.ENG_SC_C # L.TO, p-value 0.005). These significant variables and moderations have been grouped and tested simultaneously in the model (2.6): none of them remains significant.

Before assessing the support to the hypotheses, a consideration has to be made. The analyses of models (2) and (2.1) present opposing results: the former, a negative relation between incentives and savings, the latter a positive relation. These two models evaluated two different samples: the former having 568 observation and the latter 1285. For the sake of presenting a final result, model (2) prevails on model (2.1) as its goodness of fit is better ($R\text{-square}_{(2)}=0.115$; $R\text{-square}_{(2.1)}=0.044$).

These results support H1, claiming an effect of supply chain partner engagement on the performance, partially support H2, claiming a positive effect of incentives on the performance, and partially support H3, claiming an association between opportunities identification and more favorable performance. These models do not show any support to H3.

Table 6: Panel regression results (dependent variable: Emissions)

Emissions	(1.1) [b]	(1.2) [b]	(1.3) [b]	(1.4) [b]	(1.5) [b]	(1.6) [b]	(1.7) [b]
EMPL	**-0.295***	-0.223	-0.239	-0.241	-0.215	-0.224	-0.240
2.NC	**0.088****	**0.109***	**0.104***	**0.111***	**0.109***	**0.112***	**0.094**
3.NC	0.068	**0.104***	**0.103***	**0.107***	**0.102***	**0.103***	**0.100***
4.NC	**0.070***	**0.088****	**0.076****	**0.080****	**0.071***	**0.082****	**0.079****
2.SIC_M_NM	**-0.038***	**-0.051****	**-0.043***	**-0.060****	**-0.056****	**-0.058****	**-0.047***
ENG_SC_S	**0.318****						0.084
CG_INC	0.076						**-0.135***
ENG_SC_S # CG_INC	**-0.295***						0.041
L.CG_INC				0.016		0.019	
L.ENG_SC_S		**0.193****		-0.106		-0.038	**0.227****
L.TR		0.055	-0.098				-0.021
L.ENG_SC_S # L.TR		**-0.087****					**-0.098****
L.ENG_SC_C			**-0.292***		-0.283		**-0.313****
L.ENG_SC_C # L.TR			**0.122***				**0.129****
L.ENG_SC_S # L.CG_INC				0.094		0.104	
L.TO					-0.129	-0.020	
L.ENG_SC_C # L.TO					0.140		
L.ENG_SC_S # L.TO						-0.033	
Number of observations	2241	1450	1450	1466	1454	1454	1450
R-Square	0.024	0.019	0.025	0.016	0.026	0.019	0.035

p-values in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Table 7: Panel regression results (dependent variable: Savings)

Savings	(2.1) [b]	(2.2) [b]	(2.3) [b]	(2.4) [b]	(2.5) [b]	(2.6) [b]	(2.7) [b]
EMPL	1.613	0.140	0.258	0.906	0.269	0.894	0.126
2.NC	**1.686*****	**0.974***	**0.941***	0.905	0.854	0.902	**1.015***
3.NC	**1.565*****	**1.620*****	**1.487****	**1.665*****	**1.498****	**1.679*****	**1.623****
4.NC	**1.868*****	**1.326****	**1.369****	**1.421****	**1.435****	**1.440****	**1.426****
2.SIC_M_NM	**0.836****	**0.823***	**0.853***	0.696	**0.876***	0.759	**0.837***
ENG_SC_S	-0.214						-1.425
CG_INC	**-2.109****						-1.659
ENG_SC_S # CG_INC	0.598						1.400
L.CG_INC				0.372		0.482	
L.ENG_SC_S		0.347		**4.616***		3.442	0.185
L.TR		-0.237	0.536				0.200
L.ENG_SC_S # L.TR		0.326					0.403
L.ENG_SC_C			1.310		**2.045****		1.353
L.ENG_SC_C # L.TR			-0.761*				-0.775*
L.ENG_SC_S # L.CG_INC				**-3.798***		-4.123*	
L.TO					0.523	-0.559	
L.ENG_SC_C # L.TO					**-1.205*****		
L.ENG_SC_S # L.TO						0.729	
Number of observations	1285	752	752	761	756	756	752
R-Square	0.044	0.036	0.037	0.050	0.043	0.053	0.045

p-values in parentheses
* p<0.10, ** p<0.05, *** p<0.01

The control variables showed that both the proxies for the *firm's dimension* observed are significant on both the performances. Larger firms by the *number of employees* are related to lower *emission* levels, while more *globally extended* firms are related to higher *emissions* and higher *savings*. Finally, the distinction between manufacturers and not-manufactures is significant: manufacturers (2.SIC_NM_M) have lower emission levels and generate more savings than non-manufactures.

DISCUSSION

The first hypothesis (H1) aimed to understand whether engaging partners along the supply chain is reflected in improved performances. Collaboration with *customers* is directly associated with improved c*arbon performance* while interacting with the identified risks worsens performance. Furthermore, collaboration with *suppliers* directly causes worse *carbon performance*. It is not possible to establish which effect is predominant due to the transformation of the dependent variable; therefore, this study can only allege the existence of an effect of supply chain partner engagement on carbon performance. Further investigations will cover this matter. These negative results can be explained by the fact that firms are strongly encouraged to participate in the CDP program by their buyers, who could establish a selection criterion based on participation in such programs (Villena & Dhanorkar, 2020); firms yield to normative pressure and comply with the buyers' request. However, it does not mean they become fully engaged in sustainability. Jira and Toffel (2013) studying such a phenomenon, alleging that it is clear that many firms disclose through the CDP questionnaire in what they call a "symbolic" disclosure, as these firms voluntarily avoid disclosing potentially harmful information.

Conversely, both *supplier* and *customer* collaboration are directly associated with improved *environmental savings* and, again, *customer* collaboration interacts with *opportunities management* to worsen the performance. Therefore, this study can allege a positive influence of *supply chain partners engagement* on *environmental monetary performance*. This result appears relatively intuitive that collaborating with value chain partners would reduce the load of investments on one element to spread it among the partners, thereby generating more savings. Comparing the results on the environmental performance and the environmental monetary performance, the results seem to be controversial. While they share the same interactive worsening effect with *customer collaboration* on *managerial incentives*, the direct effects on the financial performance are positive, and those on carbon performance are not; instead, they are inconclusive. This could signify that collaboration on climate-change-related issues has not yet reached a best practice level, as the majority still hardly manages to step out "business-as-usual". Having established these relations, the obtained results support the literature claiming that supply chain partners engagement stimulates carbon disclosure (Jira & Toffel, 2013; Hahn et al., 2015), which is stimulated by less favorable carbon performances (Luo, 2019).

The second hypothesis (H2) aimed at understanding whether providing *incentives to manage climate-change-related issues* was enough to improve the performances. Providing *incentives* directly improves both the performances, while the interaction with *supplier engagement* improves *emissions* but worsens *savings*. This improving effect is in line with the extant literature, confirming the results that have already been obtained (Dahlmann et al., 2017; Luo & Tang, 2020). The interactive negative effect, which contrasts the extant literature, will be further investigated. Since providing incentives is a proxy of a developed strategy to manage sustainability, and it appears that they are well received as they manage to stir better performances by effectively aligning the different objectives of the manager and the organization (Matt & George, 1998; Tosi & Greckhamer, 2004; Bruce et al., 2005; Gerhart et al., 2009), it is possible that the sample analyzed is constituted of generally sustainably behaving companies, or firms employing "clever accounters" (Givetash, 2019).

The third hypothesis (H3) aimed at understanding whether the identification of climate-change-related risks negatively affected the performances. The results show partial support to the results obtained by Elijido-Ten (2017), as the total number of *risks* has both a worsening direct and interactive effect on and improving interactive effect on *carbon performance*.

The fourth hypothesis (H4) aimed at understanding whether the identification of *climate-change-related opportunities* positively affected the performances. The results showed that *opportunities* have an indirect worsening effect on *savings*. The results contrasting with the literature can be attributed to the different samples (different channels through which CDP gathers participants, for whom the participation is reflected in a different cash flow stream – revenue or investments, or revenue and investments, or neither) and different time frames.

CONCLUSION

The present study had observed corporate sustainability, illustrating how its characteristics concerning the implementation of Sustainable Supply Chain Management practices impact the firm's environmental performances with regards to the Carbon Disclosure Project.
Concerning *supply chain collaboration*, one would expect that it should improve carbon performance. Conversely, this has not been entirely confirmed by this study. Though value chain partners can indeed stir each other towards better transparency, this does not imply that the environmental performances are supposed to improve. The results outline the need to better align climate change-related targets along the supply chain to achieve consistent and meaningful results, such as carbon abetment.
All the observed drivers of Sustainable Supply Chain Management (value chain partner engagement, corporate governance and risk management) do not have a stable positive or negative effect on the performances; this could be attributed to the fact that even though the companies in the sample seem to be generally well sustainably behaving, they might not achieve best practice nor shattering "business-as-usual". Thereby is highlighted the need to care and do more to set and achieve more ambitious targets if the goals regarding carbon emission abatement (Net Zero Emissions by 2050) want to be met.
This study's limitations can be found first in the short-term analysis (i.e., with one year lag at maximum); therefore, a long-term pattern could not be defined. Furthermore, though being related exclusively to climate change practices, the savings considered as performance are a financial performance. Finally, supply chain engagement has been considered a binary variable defining whether there was partner engagement. A limitation could be found in this type of variable. Another possible limitation lies in the sample structure: a large part of it participated only once and did not necessarily answer all the questions that CDP provided them. This behavior could have also been adopted by those firms that participated for all the six years observed. To address these limitations, future developments could investigate the tested relations with different time lags to seek for a time pattern behavior, evaluate other aspects of environmental performance, evaluate different measures as proxies of supply chain engagement to determine more conclusive results, and investigate whether if the commitment to the CDP-SCP or transparency are highlighted in a particular cluster of the sample.
Moreover, adding non-listed firms to the sample, thus exploiting all the data that CDP gathers, could bring significant results. This could be achieved by implementing the CDP dataset with data from other databases. Other meaningful results could come from investigating the difference between CDP partners, namely the buyers, and non-partners, namely the suppliers.

REFERENCES

Agarwal, A., Giraud-Carrier, F. C., & Li, Y. (2018). A mediation model of green supply chain management adoption: The role of internal impetus. *International journal of production economics*, 205, 342-358.

Ayuso, S., Roca, M., & Colomé, R. (2013). SMEs as "transmitters" of CSR requirements in the supply chain. *Supply chain management.*, 18(5), 497-508.

Belsley, D. A. (1980). *Regression diagnostics : identifying influential data and sources of collinearity / David A. Belsley, Edwin Kuh, Roy E. Welsch*. New York [etc.: New York etc. : Wiley.

Ben-Amar, W., & Mcilkenny, P. (2015). Board Effectiveness and the Voluntary Disclosure of Climate Change Information. *Business Strategy and the Environment*, 24(8), 704-719.

Benn, S., Edwards, M., & Williams, T. (2018). Organizational change for corporate sustainability.

Blanco, C., Caro, F., & Corbett, C. J. (2016). The state of supply chain carbon footprinting: analysis of CDP disclosures by US firms. *Journal of cleaner production*, 135, 1189-1197.

Bonini, S., & Swartz, S. (2014). Profits with purpose: How organizing for sustainability can benefit the bottom line. *McKinsey on Sustainability & Resource Productivity*, 2(1), 1-15.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287-1294.

Bruce, A., Buck, T., & Main, B. G. M. (2005). Top Executive Remuneration: A View from Europe. *Journal of management studies*, 42(7), 1493-1506.

Chan, M. C., Watson, J., & Woodliff, D. (2014). Corporate Governance Quality and CSR Disclosures. *Journal of Business Ethics*, 125(1), 59-73.

Cho, S. Y., Lee, C., & Pfeiffer, R. J. (2013). Corporate social responsibility performance and information asymmetry. *Journal of Accounting and Public Policy*, 32(1), 71-83.

Cook, R. D., & Weisberg, S. (1983). Diagnostics for Heteroscedasticity in Regression. *Biometrika*, 70(1), 1-10.

Dahlmann, F., Branicki, L., & Brammer, S. (2017). 'Carrots for Corporate Sustainability': Impacts of Incentive Inclusiveness and Variety on Environmental Performance. *Business Strategy and the Environment*, 26(8), 1110-1131.

Damert, M., Feng, Y., Zhu, Q., & Baumgartner, R. J. (2018). Motivating low-carbon initiatives among suppliers: The role of risk and opportunity perception. *Resources, Conservation and Recycling*, 136, 276-286.

Depoers, F., Jeanjean, T., & Jérôme, T. (2016). Voluntary Disclosure of Greenhouse Gas Emissions: Contrasting the Carbon Disclosure Project and Corporate Reports. *Journal of Business Ethics*, 134(3), 445-461.

Dickey, D. A., & Fuller, W. A. (1979). Dsitribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Associations*, 3(2), 1-10.

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American sociological review*, 48(2), 147-160.

Dobler, M. (2008). Incentives for risk reporting — A discretionary disclosure and cheap talk approach. *The international journal of accounting : education and research*, 43(2), 184-206.

Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal*, 3(2), 168-177.

Elijido-Ten, E. O. (2017). Does recognition of climate change related risks and opportunities determine sustainability performance? *Journal of Cleaner Production*, 141, 956-966.

Elkington, J. (1994). *Cannibals with forks: The Triple Bottom Line of 21st Century Business*. Oxford: Capstone.

Elkington, J. (2018). 25 years ago I coined the phrase “triple bottom line.” Here’s why it’s time to rethink it. *Harvard Business Review*, 25, 2-5.

Fabrizio, K. R., & Kim, E.-H. (2019). Reluctant Disclosure and Transparency: Evidence from Environmental Disclosures. *Organization Science*, 30(6), 1207-1231.

Farooque, M., Zhang, A., Thürer, M., Qu, T., & Huisingh, D. (2019). Circular supply chain management: A definition and structured literature review. *Journal of Cleaner Production*, 228, 882-900.

Fiegenbaum, A., & Thomas, H. (1988). Attitudes toward Risk and the Risk-Return Paradox: Prospect Theory Explanations. *Academy of Management journal*, 31(1), 85-106.

Freeman, R. E. (1984). *Strategic management : a stakeholder approach / R. Edward Freeman*. Boston [etc.: Boston etc. : Pitman.

Friedman, M. (1970). The social responsibility of business is to increase its profits. *New York Times Magazine*, 122(126), 32-33.

Gallego-Álvarez, I., García-Sánchez, I. M., & Da Silva Vieira, C. (2014). Climate Change and Financial Performance in Times of Crisis. *Business Strategy and the Environment*, 23(6), 361-374.

Gerhart, B., Rynes, S. L., & Fulmer, I. S. (2009). Pay and performance: individuals, groups, and executives. *Academy of Management annals*, 3(1), 251-315.

Giannarakis, G., Zafeiriou, E., & Sariannidis, N. (2017). The Impact of Carbon Performance on Climate Change Disclosure. *Business Strategy and the Environment*, 26(8), 1078-1094.

Giunipero, L. C., Hooker, R. E., & Denslow, D. (2012). Purchasing and supply management sustainability: Drivers and barriers. *Journal of Purchasing and Supply Management*, 18(4), 258-269.

Givetash, L. (2019).
Greta Thunberg blasts soft emissions targets at U.N. climate summit. *NBC News*.

González-Benito, J., & González-Benito, Ó. (2008). A study of determinant factors of stakeholder environmental pressure perceived by industrial companies. *Business strategy and the environment.*, n/a-n/a.

Greene, W. H. (2003). *Econometric Analysis, 5th edn Prentice-Hall*.

Gualandris, J., & Kalchschmidt, M. (2014). Customer pressure and innovativeness: Their role in sustainable supply chain management. *Journal of Purchasing and Supply Management*, 20(2), 92-103.

Guenther, E., Guenther, T., Schiemann, F., & Weber, G. (2016). Stakeholder Relevance for Reporting. *Business & Society*, 55(3), 361-397.

Hahn, R., Reimsbach, D., & Schiemann, F. (2015). Organizations, Climate Change, and Transparency. *Organization & Environment*, 28(1), 80-102.

Hausman, J. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251-71.

Haverkamp, D. J., Bremmers, H., & Omta, O. (2010). Stimulating environmental management performance. *British food journal.*, 112(11), 1237-1251.

Henriques, I., & Sadorsky, P. (1999). The relationship between environmental commitment and managerial perceptions of stakeholder importance. *The Academy of Management journal.*, 42(1), 87-99.

Hsu, C. C., Choon Tan, K., Hanim Mohamad Zailani, S., & Jayaraman, V. (2013). Supply chain drivers that foster the development of green initiatives in an emerging economy. *International Journal of Operations & Production Management*, 33(6), 656-688.

Hsueh, L. (2019). Opening up the firm: What explains participation and effort in voluntary carbon disclosure by global businesses? An analysis of internal firm factors and dynamics. *Business Strategy and the Environment*, 28(7), 1302-1322.

Huang, Y. A., Weber, C. L., & Matthews, H. S. (2009). Categorization of Scope 3 Emissions for Streamlined Enterprise Carbon Footprinting. *Environmental Science & Technology*, 43(22), 8509-8515.

Jackson, S. E., & Dutton, J. E. (1988). Discerning Threats and Opportunities. *Administrative Science Quarterly*, 33(3), 370.

Jira, C., & Toffel, M. W. (2013). Engaging Supply Chains in Climate Change. *Manufacturing & service operations management*, 15(4), 559-577.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263.

Kahneman, D., Tversky, A., MacLean, L. C., & Ziemba, W. T. (2000). Choices, Values, and Frames. *Handbook of the Fundamentals of Financial Decision Making*, 4, 269-278.

Kiron, D., Kruschwitz, N., Haanaes, K., & Velken, I. V. S. (2012). Sustainability nears a tipping point. *MIT Sloan Management Review* 53(2), 69.

Li, D., Huang, M., Ren, S., Chen, X., & Ning, L. (2018). Environmental Legitimacy, Green Innovation, and Corporate Carbon Disclosure: Evidence from CDP China 100. *Journal of Business Ethics*, 150(4), 1089-1104.

Luo, L. (2019). The influence of institutional contexts on the relationship between voluntary carbon disclosure and carbon emission performance. *Accounting & Finance*, 59(2), 1235-1264.

Luo, L., Lan, Y.-C., & Tang, Q. (2012). Corporate Incentives to Disclose Carbon Information: Evidence from the CDP Global 500 Report. *Journal of International Financial Management & Accounting*, 23(2), 93-120.

Luo, L., & Tang, Q. (2020). Corporate governance and carbon performance: role of carbon strategy and awareness of climate risk. *Accounting & Finance*.

Luo, L., Tang, Q., & Lan, Y. C. (2013). Comparison of propensity for carbon disclosure between developing and developed countries: A resource constraint perspective. *Accounting research journal*, 26(1), 6-34.

Luo, L., Tang, Q., & Peng, J. (2018). The direct and moderating effects of power distance on carbon transparency: An international investigation of cultural value and corporate social responsibility. *Business Strategy and the Environment*, 27(8), 1546-1557.

Mahapatra, S. K., Schoenherr, T., & Jayaram, J. (2021). An assessment of factors contributing to firms' carbon footprint reduction efforts. *International Journal of Production Economics*, 235, 108073.

Mateo-Márquez, A. J., González-González, J. M., & Zamora-Ramírez, C. (2019). Countries' regulatory context and voluntary carbon disclosures. *Sustainability accounting, management and policy journal (Print)*, 11(2), 383-408.

Matt, B., & George, T. M. (1998). Relationships among Risk, Incentive Pay, and Organizational Performance. *Academy of Management journal*, 41(3), 283-297.

Ott, C., Schiemann, F., & Günther, T. (2017). Disentangling the determinants of the response and the publication decisions: The case of the Carbon Disclosure Project. *Journal of Accounting and Public Policy*, 36(1), 14-33.

Pagell, M., & Wu, Z. (2009). BUILDING A MORE COMPLETE THEORY OF SUSTAINABLE SUPPLY CHAIN MANAGEMENT USING CASE STUDIES OF 10 EXEMPLARS. *Journal of Supply Chain Management*, 45(2), 37-56.

Saeed, M., & Kersten, W. (2019). Drivers of Sustainable Supply Chain Management: Identification and Classification. *Sustainability*, 11(4), 1137.

Schrettle, S., Hinz, A., Scherrer -Rathje, M., & Friedli, T. (2014). Turning sustainability into action: Explaining firms' sustainability efforts and their impact on firm performance. *International Journal of Production Economics*, 147, 73-84.

Serafeim, G. (2020). Social-impact efforts that create real value. *Harvard Business Review*, 98(5), 37-48.

Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of cleaner production*, 16(15), 1699-1710.

Stanny, E. (2013). Voluntary Disclosures of Emissions by US Firms. *Business Strategy and the Environment*, 22(3), 145-158.

SustainAbility (2012). Rate the Raters 2012: Polling the Experts. *Rate the Raters*.

SustainAbility (2020). Rate the Raters 2020: Investor Survey and Interview Results. *Rate the Raters*.

Tate, W. L., Ellram, L. M., & Kirchoff, J. F. (2010). CORPORATE SOCIAL RESPONSIBILITY REPORTS: A THEMATIC ANALYSIS RELATED TO SUPPLY CHAIN MANAGEMENT. *Journal of Supply Chain Management*, 46(1), 19-44.

Tay, M. Y., Abd Rahman, A., Aziz, Y. A., Sidek, S. J. I. J. o. S. S., & Humanity (2015). A review on drivers and barriers towards sustainable supply chain practices. 5(10), 892.

Tosi, H. L., & Greckhamer, T. (2004). Culture and CEO Compensation. *Organization science (Providence, R.I.)*, 15(6), 657-670.

Tsai, F. M., Bui, T.-D., Tseng, M.-L., Ali, M. H., Lim, M. K., & Chiu, A. S. F. (2021). Sustainable supply chain management trends in world regions: A data-driven analysis. *Resources, Conservation and Recycling*, 167, 105421.

Verbeek, M. (2017). *A guide to modern econometrics / Marno Verbeek (5th edition. ed.)*. Hoboken: Hoboken : Wiley Custom.

Villena, V. H., & Dhanorkar, S. (2020). How institutional pressures and managerial incentives elicit carbon transparency in global supply chains. *Journal of operations management.*, 66(6), 697-734.

Walker, H., Di Sisto, L., & McBain, D. (2008). Drivers and barriers to environmental supply chain management practices: Lessons from the public and private sectors. *Journal of Purchasing and Supply Management*, 14(1), 69-85.

Whelan, T., & Fink, C. (2016). The Comprehensive Business Case for Sustainability. *Harvard Business Review*, 21, 2-8.

Wilhelm, M. M., Blome, C., Bhakoo, V., & Paulraj, A. (2016). Sustainability in multi-tier supply chains: Understanding the double agency role of the first-tier supplier. *Journal of Operations Management*, 41(1), 42-60.

Zhu, Q., & Sarkis, J. (2007). The moderating effects of institutional pressures on emergent green supply chain practices and performance. *International journal of production research*, 45(18-19), 4333-4355.

Zhu, Q., Sarkis, J., & Lai, K.-H. (2007). Green supply chain management: pressures, practices and performance within the Chinese automobile industry. *Journal of Cleaner Production*, 15(11-12), 1041-1052.

DECISION SCIENCES INSTITUTE

Blockchain Impacts on Global Supply Chain Efficiency and Effectiveness Performance

Pedro M. Reyes
Baylor University
Email: pedro_reyes@baylor.edu

Michael Gravier
Bryant University
Email: mgravier@bryant.edu

Patrick Jaska
University of Mary Hardin-Baylor
Email: pjaska@umhb.edu

John K. Visich
Bryant University
Email: jvisich@bryant.edu

ABSTRACT

Since the term "blockchain" was introduced in 2008, it has received attention from practitioners and academics as an emerging technology. Much like the introductory of other emerging technologies, there are disruptions and hype regarding its impact on global supply chain performance. This paper examines how blockchain technology impacts global supply chain efficiency and effectiveness performance. We utilized the Mooney *et al.* (1996) framework to organize this evidence into operational and managerial business value processes, and the effect: automational, informational, and transformational. In addition, we have identified possible reasons why blockchain is not following the Mooney *et al.* (1996) stage model.

KEYWORDS: Blockchain, Supply Chain, Performance

DECISION SCIENCES INSTITUTE

Integrated Operations Planning with Penalty Costs

Gang Wang
University of Massachusetts Dartmouth
Email: gwang1@umassd.edu

ABSTRACT

This paper considers an integrated planning problem with delivery due dates and penalty costs over a capacitated supply and distribution network, which consists of contracted suppliers, processing centers (PCs), and demand points (DPs). Shipped from suppliers to PCs, semi-finished products are further processed and/ or require custom configuration. The problem is to assign suppliers and DPs to PCs, and schedule production and distribution operations optimally, such that the total shipping costs plus penalty costs are minimized, subject to the capacity of suppliers and PCs and delivery due dates.

KEYWORDS: Integrated planning; delivery due date; service level; penalty cost; mixed integer nonlinear programming.

INTRODUCTION

With supply chains becoming increasingly complex, companies are struggling to cope with a wide range of challenges such as the management of cost and customer relationship, especially more demanding customer requirements (e.g., timely and perfect-order delivery) according to the Global Supply Chain survey by PwC's (2013). To achieve a fast and efficient supply chains, the need for the tighter supply chain integration in the context of complex global operations has become even more vital.

Our study on integrated distribution planning with delivery deadlines was motivated by the global supply chain of a manufacturer, who produces time-sensitive products such as fast fashion brands (e.g., Zara and Forever 21) or high-tech products that typically have short life cycle and are sold in a very short selling season. Because of high demand uncertainty of the products, retailers typically do not place orders until reliable market information is available shortly before a selling season. On the other hand, because there are significant markdowns for unsold products at the end of the selling season, the manufacturer runs a high risk if it starts production processing early before it receives orders from the retailers. As a result, the manufacturer would not start production until orders from the retailers have been placed shortly before the selling season. Since the time-sensitive products cannot be inventoried due to their short life cycle the delivery must be made to customers as soon as possible at a low cost. Lead times and efficiencies play an extremely important role in gaining competitive advantage. Therefore, the supply and distribution operations must be highly integrated, **to** speed up the supply chain, meet a customer's delivery deadlines and provide a guaranteed level of customer service. On the other hand, poor planning would cause product expiration and cancellation of orders, which in in turn lead to a poor reputation and a decline in demand.

MODEL

In this section, a capacitated supply and distribution network is considered, where raw materials, or semi-finished products are first shipped from a set $\mathcal{S} = \{1,2,\cdots,S\}$ of contracted suppliers, on as-needed basis, to a number $\mathcal{J} = \{1,2,\cdots,J\}$ of demand points (DPs) through a manufacturer, who geographically owns several processing centers (PCs), $\mathcal{N} = \{1,2,\cdots,N\}$, each with a capacity of C_n and the processing time τ_n per unit. DPs request order quantity Q_j, and delivery deadline D_j, $j \in \mathcal{J}$. After receiving all DPs' order over a period, the manufacturer must make decisions on the acceptance and assignment of orders to the PCs due to the constraints on its capacity and order delivery deadlines, then deciding on suppliers' delivery due date t_{sn}^{sup}, $s \in \mathcal{S}, and\ n \in \mathcal{N}$. To maintain target service level, if some supplier cannot meet the manufacturer's quoted due date, local subcontractors will be employed. For any shipments from suppliers to PCs, restricted by the capacity of W_s, suppliers charge a shipping rate of a_{sn}, $s \in \mathcal{S}, and\ n \in \mathcal{N}$. Upon the receipt of all required semi-finished products, the PCs start processing according to DP's custom requirements. Once the production at the PCs is completed, finished products are immediately distributed to the DPs without inventory. As specified in agreed-on contract with the DPs, the manufacturer is imposed a penalty cost for any unfulfilled order, denoted by p_j, $j \in \mathcal{J}$. Each shipment between the PCs and DPs incurs the fixed shipping cost f_j and shipping rate v_j during the transit time t_{nj}^{dp}, $j \in \mathcal{J}$. The problem is to determine which set of orders should be accepted and assigned to the PCs, suppliers' delivery due date and operations schedules for processing at the PCs and delivery of finished products to the DPs.

Model Formulation

To facilitate the analysis of this problem, let us define decision variables and summarize notations in the following.

Decision Variables

x_{sn} = Shipping quantity from supplier s to $\mathrm{PC_n}$, $s \in \mathcal{S}$, $n \in \mathcal{N}$;

t_{sn} = Delivery due date of supplier s to $\mathrm{PC_n}$, $s \in \mathcal{S}$, $n \in \mathcal{N}$;

y_{nj} = $\begin{cases} 1, \text{if } \mathrm{DP_j}'\text{s order is accepted and assigned to } \mathrm{PC_n} \\ 0, \text{otherwise} \end{cases}$, $n \in \mathcal{N}$, $j \in \mathcal{J}$.

Index Set

$\mathcal{S}$ = Set of contracted suppliers, i.e., $\mathcal{S} = \{1,2,\cdots,S\}$;

$\mathcal{J}$ = Set of demand points, i.e., $\mathcal{J} = \{1,2,\cdots,J\}$;

$\mathcal{N}$ = Set of processing centers, i.e., $\mathcal{N} = \{1,2,\cdots,N\}$;

Parameters

W_s = Supply capacity of supplier s, $s \in \mathcal{S}$;

C_n = Processing capacity of $\mathrm{PC_n}$, $n \in \mathcal{N}$;

Q_j = Order quantity at the $\mathrm{DP_j}$, $j \in \mathcal{J}$;

D_j = Delivery deadline specified by $\mathrm{DP_j}$, $j \in \mathcal{J}$;

T_{sn}^{sup} = Earliest delivery date from supplier s to $\mathrm{PC_n}$, $s \in \mathcal{S}, n \in \mathcal{N}$;

t_{nj}^{dp} = Shipping time from $\mathrm{PC_n}$to $\mathrm{DP_j}$, $n \in \mathcal{N}, j \in \mathcal{J}$;

L	=	Sufficiently large number.
a_{sn}	=	Shipping rate from supplier s to $\mathrm{PC_n}$, $s \in \mathcal{S}$, $n \in \mathcal{N}$;
μ_n	=	Processing time at the $\mathrm{PC_n}$ per unit, $n \in \mathcal{N}$;
v_{nj}	=	Shipping rate from $\mathrm{PC_n}$to $\mathrm{DP_j}$, $n \in \mathcal{N}, j \in \mathcal{J}$;
f_{nj}	=	Fixed shipping cost from $\mathrm{PC_n}$to $\mathrm{DP_j}$, $n \in \mathcal{N}, j \in \mathcal{J}$;
p_j	=	Manufacturer's penalty cost for not fulfilling $\mathrm{DP_j}$'s order, $j \in \mathcal{J}$;
α	=	Manufacturer's target service level.

Objective function

$$Min \quad \sum_{s \in \mathcal{S}} \sum_{n \in \mathcal{N}} a_{sn} x_{sn} + \sum_{n \in \mathcal{N}} \sum_{j \in \mathcal{J}} (v_{nj} Q_j + f_{nj}) y_{nj} + P(y) \tag{0.1}$$

Constraints

- Constraints on suppliers' capacity

$$\sum_{n \in \mathcal{N}} x_{sn} \leq W_s,\ s \in \mathcal{S} \tag{0.2}$$

- Constraints on PCs' processing capacity

$$\sum_{j \in \mathcal{J}} Q_j y_{nj} \leq C_n,\ n \in \mathcal{N} \tag{0.3}$$

- Constraints on manufacturer's delivery deadlines

$$D_j \geq t_{sn} + \sum_{j \in \mathcal{J}} \left[\mu_n \left(Q_j y_{nj} \right) \right] + t_{nj}^{dp}\, y_{nj} - L \left(1 - y_{nj} \right),\ s \in \mathcal{S}, n \in \mathcal{N}, j \in \mathcal{J} \tag{0.4}$$

- Flow balance through PCs

$$\sum_{s \in \mathcal{S}} x_{sn} = \sum_{j \in \mathcal{J}} Q_j y_{nj},\ n \in \mathcal{N} \tag{0.5}$$

- Constraints on DPs' sourcing strategy

$$\sum_{n \in \mathcal{N}} y_{nj} \leq 1,\ j \in \mathcal{J} \tag{0.6}$$

- Integrality and non-negativity

$$x_{sn} \geq 0,\ t_{sn} \geq T_{sn}^{sup},\ \ y_{nj} \in \{0,1\},\ s \in \mathcal{S},\ n \in \mathcal{N}, j \in \mathcal{J}. \tag{0.7}$$

The first term in the objective function of (2.1) defines the shipping cost from suppliers to PCs, where $a_{sn} x_{sn}$ estimates the cost of shipping a quantity of x_{sn} from supplier s to PC_n. The second term represents the variable and fixed shipping cost from PCs to demand points. The third term denotes the total penalty costs incurred by unsatisfied demands. Constraints (2.2) ensure that all the quantities shipped out of supplier s do not exceed supplier's capacity. Constraints (2.3) ensure that PC capacities are not violated. Constraints (2.4) impose the requirement that the shipment arrives at demand point j must be no later than the specified due date. Constraints (2.5) are flow conservation constraints for each PC. Constraints (2.6) refer to the *single-sourcing requirement* that each demand point is supplied by at most one PC in accordance with the assumptions.

Model Analysis

This section contains the tighter formulation, NP-hardness proof, and polynomial solvable cases of the problem under study. The following theorem provides a valid inequality of the above constraint (2.4) to strengthen the formulation of IPC.

Theorem 1.1 There exists a valid inequality

$$\sum_{j\in\mathcal{J}} \frac{\mu_n Q_j + t_{nj}^{dp}}{D_j} y_{nj} \leq 1,\ n \in \mathcal{N} \qquad (0.8)$$

for the constraint (2.4).

Proof. To show that the constraints (2.8) is valid for the problem (2.1), we shall define three subsets of $\mathcal{J}$ in the following

$$\mathcal{J}_1 = \left\{ \mathrm{J} \in \mathcal{J} \middle| \frac{\mu_n Q_j + t_{nj}^{dp}}{D_j} > 1 \right\},$$

$$\mathcal{J}_2 = \left\{ \mathrm{J} \in \mathcal{J} \middle| \frac{\mu_n Q_j + t_{nj}^{dp}}{D_j} = 1 \right\},$$

$$\mathcal{J}_3 = \left\{ \mathrm{J} \in \mathcal{J} \middle| \frac{\mu_n Q_j + t_{nj}^{dp}}{D_j} < 1 \right\}.$$

It is evident that $\mathrm{y_{nj}} = 0,\ \forall\ \mathrm{j} \in \mathcal{J}_1$, $\mathrm{y_{nj}} = 1,\ \forall\ \mathrm{j} \in \mathcal{J}_2$, and $\mathrm{y_{nj}} = 0,\ \forall\ \mathrm{j} \in \mathcal{J} \backslash \mathcal{J}_2$. Now let us turn to the subset, $\mathcal{J}_3$. We only need to show that any feasible solution of (2.1) will satisfy the constraints (2.8). Let y be the optimal order assignment to (2.1), and define $\mathcal{J}_4$ as the subset of optimally assigned customers in the problem (2.1), i.e.,

$$\mathcal{J}_4 = \{\mathrm{J} \in \mathcal{J} | y_{nj} = 1\},\ \forall n \in \mathcal{N}.$$

For any $\mathrm{k} \in \mathcal{J}_4$, we consider the constraint (2.4) and obtain the following results,

$$\sum_{\mathrm{j}\in\mathcal{J}\backslash\{k\}} (\mu_\mathrm{n} \mathrm{Q_j}) \mathrm{y_{nj}} + (\mu_\mathrm{n} \mathrm{Q_k}) \mathrm{y_{nk}} + \mathrm{t_{nk}} \mathrm{y_{nk}} \leq \mathrm{D_k}.$$

Since $\sum_{\mathrm{j}\in\mathcal{J}\backslash\{k\}} (\mu_\mathrm{n} \mathrm{Q_j}) \mathrm{y_{nj}} \geq 0$, it follows that $(\mu_\mathrm{n} \mathrm{Q_k}) \mathrm{y_{nk}} + \mathrm{t_{nk}} \mathrm{y_{nk}} \leq \mathrm{D_k}$, i.e., $\mathcal{J}_4 \subseteq \mathcal{J}_3$. □

By replacing the constraint (2.4) with the valid inequality, (2.1) can be represented as follows:

$$\text{IPC:} \quad Min \quad \sum_{s\in\mathcal{S}} \sum_{n\in\mathcal{N}} a_{sn} x_{sn} + \sum_{n\in\mathcal{N}} \sum_{j\in\mathcal{J}} (v_{nj} Q_j + f_{nj}) y_{nj} + G(y)$$
$$s.t.$$
$$(2.2),\ (2.3),\ (2.5)\ \text{-}(2.8).$$

Throughout the remaining discussion in this paper, the problem under study is denoted by **IPC**. The theorem is regarding the complexity of the problem, **IPC**.

Theorem 1.2 *IPC is strongly NP-hard.*

Proof. It suffices to show that some known NP-complete problem is reducible to IPC with a polynomial time algorithm.

3-Partition: Given integers t, b, and a set of integer a_1, a_2, $\cdots$, a_{3t} such that $\frac{b}{4} \leq a_i \leq \frac{b}{2}$, $i = 1,2,\cdots,3t$, $\sum_{i=1}^{3t} a_i = tb$, are there t pairwise disjoint three element subsets $A_j \subset \{1,2,\cdots,3t\}$ such that $\sum_{i \in A_j} a_i = b$, $j = 1,2,\cdots,t$?

CONCLUSIONS

We investigated a scheduling problem with delivery due dates in an integrated manner. In this paper, the supply chain considered consists of contracted suppliers, PCs, and DPs. We formulated this problem as a mixed-integer programming problem and provided some mathematical properties. Notably, we show that the problem is strongly NP-hard. After solving this problem, we can determine the assignments of suppliers and DPs to PCs and schedules of production and distribution operations.

REFERENCES

Pwc. Next-generation supply chains Efficient, fast and tailored. https://www.pwc.com/gx/en/consulting-services/supply-chain/global-supply-chain-survey/assets/global-supply-chain-survey-2013.pdf. Accessed by May 31, 2021.

Chua, J. M. (2019). The environment and economy are paying the price for fast fashion — but there's hope. https://www.vox.com/2019/9/12/20860620/fast-fashion-zara-hm-forever-21-boohoo-environment-cost. Access by May 31, 2021.

DECISION SCIENCES INSTITUTE
Healthcare Operations Scheduling

Gang Wang
University of Massachusetts Dartmouth
Email: gwang1@umassd.edu

ABSTRACT

Scheduling has been a vital part of healthcare operations, which concerns scheduling of tests among different automated instruments and staff. The problem is to find a scheduling policy that minimizes the total completion time with capacity constraints. This paper formulates operations scheduling problem as mixed integer program and develops a scheduling policy for tests performed on automated equipment. A solution approach is presented using linear programming and genetic algorithm. Since linear programming can find close-to-optimal solution, it is used in the selection process of the genetic algorithm to improve the possibility of finding the optimal solution.

KEYWORDS: Healthcare operations; scheduling; mixed integer programming; genetic algorithm.

INTRODUCTION

Healthcare currently accounts for 15% of the US Gross Domestic Product and is expected to reach 19% within the coming decade, according to Centers for Medicare and Office of the Actuary Medicaid Services. The increasing healthcare operations costs come from new advances in expensive treatment technologies and pharmaceuticals, unfavorable trends in population demographics such as aging, obesity and chronic disease, and legal expenses resulting from medical errors and malpractice. Faced with this environment of increasing costs, limited capacity and burgeoning demand, many hospitals are emphasizing the better tradeoff between customer service and operations efficiency. This in turn is forcing medical labs to reassess their operations and capacities, with the dual objectives of stabilizing revenue streams and improving healthcare access.

This study presents a mixed-integer program to model healthcare operations scheduling problem mentioned above and develops a solution approach which combines linear programming and genetic algorithm. Healthcare operations scheduling has received much attention in the literature. Cayirli and Gunes (2014) consider a weighted sum of waiting time, idle time, and overtime to schedule appointments. Ma et al. (2015) use simulation and optimization to investigate the assignments of cancer therapies to patients to find perfect patient-physician matching. Zacharias and Armony (2017) study the joint decision of panel size and the number of daily appointment slots by analyzing indirect and direct wait times.

The rest of the paper is organized as follows. In section 2, a mathematical model formulating this scheduling problem is proposed. Section 3 is concerned with the development of the solution approach. Finally, Section 4 concludes the study.

MODEL

Let us first summarize the notations to be used in our problem formulation.

- Index Sets

  I $Set\ of\ employees$
  J $Set\ of\ samples$

- Parameters

  p_j $time\ for\ processing\ sample\ j$
  c_i $capacity\ of\ employee\ i$
  q_{ij} $cost\ of\ processing\ sample\ j\ by\ employee\ i$

- Decision variables

$$y_{ij} = \begin{cases} 1, if\ sample\ j\ is\ tested\ by\ employee\ i \\ 0,\ \ otherwise \end{cases}$$

We now formulate the scheduling problem as a bi-criteria mixed integer program.

$$\boldsymbol{min \sum_i \sum_j q_{ij} y_{ij}}$$
$$\boldsymbol{s.t. \sum_i y_{ij} = 1, \forall\, j \in J}$$
$$\boldsymbol{\sum_j p_j y_{ij} \le c_i, \forall\, i \in I}$$
$$\boldsymbol{y_{ij} \in \{0, 1\}, \forall\, i \in I, j \in J}$$

In the objective function, the objective is to minimize the total costs incurred by processing samples by employees. Constraint 1 is regarding assignment restriction, i.e., each sample must be tested by one employee. Constraint 2 ensures that the total number of samples each employee handles cannot exceed the processing capacity of the employee. Throughout the remaining discussion in this paper, we shall denote the original problem by HOS.

ALGORITHM

Genetic algorithms (GAs) provide an intelligent heuristic for solving many types of combinatorial problems. The foundations of the GA approach were developed by Holland (2012). The GA attempts to mathematically mimic some of the adaptive processes of biological phenomena and combines survival of the fittest members in a population with a structured but randomized information exchange to form a search mechanism which does not demand assumptions about the search space. A key aspect of the GA has been the issue of robustness, enabling the method to be employed on a wide range of problems, unlike many traditional methods which either work well only on specialized problems or work at a variable level of performance across a wide range of problems. The GA uses three main operators (a) reproduction; (b) crossover; (c) mutation to improve the fitness of a set of trial solutions. For the GAP problem, fitness will be taken to mean the satisfaction of individual constraints and an algorithm will be required to converge to a solution that satisfies all constraints (or to show that no such solution exists). There is no accepted theory of GAs, but partial explanations are offered by Holland and Goldberg. For solving the generalized assignment problem an adaptation of a particular GA will be developed; this GA had success with solving the Set Covering Problem (SCP), another restricted structure integer programming problem with some similarities.

The algorithm developed in this section is an extension to the GA developed by Chu and Beasley. A solution is encoded as a chromosome that is represented as a binary string, whose value denotes the index of the bin to which item is assigned. A population is formed by a set of encoded solutions. Population size is a runtime parameter. This GA follows a steady-state reproduction process in which each iteration selects two parents to generate one child for replacement. The GA operators to be performed within each iteration are adapted from Chu and Beasley and listed below for the completeness of this paper:

Algorithm LGA

Step 1 Selection. Two parents are selected based on selection strategies such as binary tournament selection or rank-based tournament selection.

Step 2 Crossover. The simple single cut-point crossover operator randomly decides a cut point at which the first part of a parent is combined with the second part of the other to form the child solution.

Step 3 Mutation. The mutation operator is performed on the child solution obtained from the crossover operator to exchange the bin assignment of two randomly chosen items.

Step 4 Feasibility improvement. This operator looks at those over weighted bins after fitness evaluation for a reassignment that could keep the weight sum of the bin below its capacity. This is done by moving an item in an over weighted bin to an under weighted bin.

Step 5 Quality improvement. This operator looks at each item for a possible reassignment (to another bin) such that the solutions fitness value could be improved.

CONCLUSIONS

This study considered a healthcare operation scheduling problem regarding the assignments of tests and staff. Specifically, the problem found a schedule minimizing the total completion time subject to capacity constraints. We modeled this problem as a mixed-integer linear programming problem. Further, we proposed a linear programming based genetic algorithm to solve this problem. This solution approach is advantageous over the traditional genetic algorithm because solving a linear programming problem first can lead to a better feasible solution.

REFERENCES

Ma X, Saure´ A, Puterman ML, Taylor M, Tyldesley S (2015) Capacity planning and appointment scheduling for new patient oncology consults. Health Care Management Science, 19(4): 347–361.

Cayirli T, Gunes ED. (2014). Outpatient appointment scheduling in presence of seasonal walk-ins. Journal of Operational Research Society, 65(4):512–531.

Zacharias C, Armony M (2017) Joint panel sizing and appointment scheduling in outpatient care. Management Science, 63(11), 3978–3997.

DECISION SCIENCES INSTITUTE

Engage Community to Mitigate Terrorism Risk

Ashley Nchewi Gekpe
University of Houston Clear Lake
Email: GekpeA8272@UHCL.edu

Xiaojun (Gene) Shan
University of Houston Clear Lake
Email: shan@uhcl.edu

ABSTRACT

Terrorism poses a major challenge to the government whenever it surfaces in any nation. It degrades the economic, infrastructural and trading systems. Game-theoretic models have been widely applied to study the strategic interactions between the defender and attacker. In this paper, we build two subgames between the government and terrorist, and between the government and community, respectively. We identify conditions where the community equilibrium strategy is to support the government. The community could support either the government or the terrorists. Both sequential and simultaneous games are studied with different orders of moves. This paper finds first-mover advantage confirming previous research.

KEYWORDS: Business and Terrorism, Game Theory, Community Engagement, First-Mover Advantage

INTRODUCTION

In 2002, the Boko Haram group was founded by Mohammed Yusuf in Maidugri, Borno state under the government of Governor Mala Kachalla. Since then, Nigeria has been plagued by intermittent terrorist attacks executed by the Islamist group. “Boko Haram” translates from Hausa, an indigenous Nigerian language, to English as “Western Education is forbidden” or “Book is a sin”. In a self-acclaimed statement, the group also refers to themselves as “Jama'atu Ahlis Sunna Lidda'awati Wal-Jihad”, meaning “People Committed to the Propagation of the Prophet's Teachings and Jihad” (CNN, 2019). The Boko Haram group allegedly withdraws from participation in Nigerian politics as it is against the central Islam notion that the group is structured and regulated by the Sharia (Islamic law). However, it is difficult to categorize their attacks as non-political since they mostly conduct political attacks and attract the attention of the ruling government.

Under the rule of Yusuf, the group was motivated by the “liberation” of the people in the Northern States from Western education, constitutions and institutions. However, when the government intervened and killed Yusuf in 2009, a new leader, Abubakar Shekau, took his place and led the violent uprising of Boko Haram terrorism in its current form. It can be argued that a contributing factor to the increasing violence from the group came from the inauguration of President Goodluck Ebele Jonathan. Other factors could have been negligence from President Jonathan’s administration regarding the importance allocated to counter terrorism measures and elevated poverty and illiteracy levels in the Northern states, which leads to lower standard of living. From observing the strategies used and the frequency of attacks by the Boko Haram terrorist group, it can be gathered that they are a credible threat to concerned nations and should be treated as such. Based on the definition of a credible threat by Dixit and Nalebutt

(1991), credible threats contain two components: “a plan of sequenced actions and, the commitment to make the threat credible” (Shan and Zhuang, 2014).

This paper will utilize the game theoretic approach to study and analyze the decision making for three players: the government, community and terrorist with the aim of finding ways to stop the community from supporting terrorists. To the best of the authors’ knowledge, the community has not been studied using game theory for counter-terrorism. The main contributions include examination of the community as a player in counter-terrorism efforts and developing and solving a three-player game involving the government, community and terrorist.

Potential terrorists might come from the community possibly inspired by the ideology of the terrorist group. The community might play a critical role in counter-terrorism and other government-led initiatives. For instance, the community could make sure its members obeying the safety curfew set by the government for their safety during a pandemic or other emergency situation. Additionally, the community may serve as local aides to the government reporting suspicious incidents though security helplines and centers. This paper will attempt to answer the following questions:

- Under what conditions could the government prevent the community from supporting terrorists?
- How willing is the community to support the government as certain external and internal factors change?

This paper will identify model parameters which might be important for leading to equilibrium involving the decision of the community supporting governmental efforts in counter-terrorism.

Next, this study will highlight the feasibility of stopping the community from supporting terrorists. It will be significant in

- Highlighting the importance of the role of community in the mitigation of terrorism, and
- Encouraging the government to develop strategies to better encourage members of the community to support the government rather than the terrorists for the well-being of the nation.

LITERATURE REVIEW

The term “terrorism” has an array of definitions. According to Walter Laqueur, the sole properties of terrorism that are commonly accepted are that terrorism involves both violence and the threat of violence (Laqueur, 1987). The United States Department of State has defined terrorism as “politically motivated violence perpetrated against noncombatant targets by subnational groups or clandestine agents, usually intended to influence an audience” (Charles, 2002). From the definition, three criteria are used to portray the essence of the word. They are political motivation, violence against noncombatants and subnational groups or clandestine agents. Terrorism is considered “politically motivated” since the primary aim of terrorist attacks is to influence the ruling government’s decisions to favor the terrorists’ cause. The second criteria of “violence against noncombatants” states that the attacks are targeted towards civilians and/or members of the military, who are not on active duty (noncombatants). The third criteria of “subnational groups or clandestine agents” is controversial. Boko Haram attacks cannot fully be described as subnational since their attacks are both within North-Eastern Nigeria, and neighboring countries such as the Republic of Chad, Cameroun and Niger. Also, they can no longer be categorized collectively as clandestine agents except in terms of individual agents since some of their operatives are of known identity. According to the 2019 Global Terrorism Index, Nigeria ranks 3rd with a score of 8,597 indicating a very high terrorism impact on the country (Institute for Economics & Peace, 2019). Numerous perspectives arise when comparisons are made between terrorists and freedom fighters. In this context, demonstrations by the Nigerian Niger Delta Militants are compared to those by the Boko Haram terrorist group. Both are being labelled as terrorists by some schools of thought. Although it can be gathered that both groups are motivated by the pursuit of liberation, and achieve their goals through violence and forceful motives, there is a major difference between both groups. Freedom fighters are motivated to act against the ruling government when the government deprives the people of fundamental necessities that are standards for humane living conditions. On the other hand, terrorism acts are calculated acts of destruction on the people and infrastructure conducted with the sole purpose of causing deliberate harm to both the nation and the nation's population (Ochoche, 2013).

Furthermore, organized crime and terrorism share notable differences. Although both organized crime and terrorism share some similarities such as the need for money (to launch attacks), some differences exist between them. These differences include the aim of organized crime and terrorism being to yield economical profits and affect political activities, respectively (Passas, 2005). While organized crime mostly pursues anonymity, terrorism seeks media coverage of attacks, where the concerned terrorist group takes responsibility for the attack as a form of intimidation and as a method to attract new followers. Another distinguishing trait between organized crime and terrorism is that terrorist attacks are ocassionally planned to target prominent governement agencies or nationally recoginsed commemoration dates and anniversaries. The potential threat that terrorism poses to the targeted community worsens with time. According to the United Nations High Commission for Refugees (UNHCR), over 3.3 million people have been displaced since the Boko Haram terrorists' attacks heightened in 2014. Within the 3.3 million, over 2.5 million of the internally displaced persons (IDPs) are reportedly in North-Eastern Nigeria, whereas over 550,000 are reportedly in the Republic of Chad, Cameroon and Niger. Also, as of 2014, there were approximately 240,000 refugees in Nigeria, Cameroon, Chad and Niger altogether (United Nations High Commissioner for Refugees, 2001-2020).

Generally, the community plays a vital role in the mitigation and prevention of crimes and pandemics. This could be achieved through the formation of a Neighborhood Watch group. A Neighborhood Watch is a police-enabled program that encourages members of the community to help protect themselves and support the police. According to John Woodhead, the Vice President and Executive Chairman of Queensland Neighborhood Watch Association Inc., a major issue with all association branches was receiving cooperation from the local police departments since the degree of support varied in all states. It was reported that the establishment of the Neighborhood Watch program reduced the rate of burglaries and crimes in general in the Queensland, Australia, and in the other states (Woodhead, 1990). Also, members of the community formed close-knit relationships with the police as they worked together towards achieving common goals. However, recently the Neigbourhood Watch seems not as effective. According to Jenny Fleming (2005), the Neigbourhood Watch despite having large numbers of volunteers are only slightly reducing crimes in areas with low crime rates. However, it could be due to the fact that the studied areas already have low-crime rates. The role of communities in providing their residents with safety and security could not be sufficiently emphasized. Analogy could be drawn between community-based counter-terrorism efforts with community-policing initiatives. Whereas there is a continuous debate of the effectiveness of centralized vs. decentralized (i.e., community-based policing) (Dragos and Boettke 2009), Boettke et al. (2016) argue that the failure of some community-policing initiatives could be due to the operational difficulties created by the strategic interactions among federal, state and local government/law enforcement agencies. Boettke has been instrumental in proposing community-oriented model for analyzing public services including policing (Boettke et al. 2013, Vlad 2016). Neighborhood Watch to fight terrorism is not a new concept. Escalante (2020) compares Night Wathers (vigilante's patrol similar to Neighborhood Watch) in the northern and southern Peruvian communities in combating terrorism. Further examining the role of the community in the fight against terrorism in Nigeria leads us to the events that occurred on May 13, 2014. Boko

Haram terrorists, who attacked three villages in Maiduguri, Borno state, were met by resistance from the community leading to the deaths of over 200 Boko Haram terrorists (CNN, 2019). This particular incident leads supports to the perceptive of Goodman (2018) that polycentric justice.

The important role of the community in detering adverse activities in their nation cannot be overstated. While highlighting the importance of the community's role in the prevention of global issues, we examine community responses and roles in deterring a recent global pandemic, the spread of the COVID-19 diseade, social distancing, increased personal and household hygiene, observing self-quarantine, obeying the federal quarantine orders, and obeying the orders put in place by essential facilities are some of the ways the community helped reduce the coronavirus spread (Harvard Health Publishing, 2020). Another highly important role that members of the community play is in supporting one another in crisis. Having support is fundamental for members of affected communities especially for children and young adults, who may not fully comprehend an on-going crisis. Community support in crisis can be discounted supplies in neighborhood stores, aiding neighbors whenever needed, and keeping one another informed on the new changes or policies (Boudreau, 2020).

The government mainly focuses on deterrence as a form of terrorism mitigation. They deter the attacks by instilling doubt and fear within the nationals regarding terrorism and the punishments for acts of terrorism if prosecuted. However, deterrence has not always been effective. In fact, it discourages innocent members of the community from providing tips as they fear being punished as severely as the terrorists if they are assumed to be members of the terrorist groups. Richardson et al. (2007) offered three strategies to terrorism mitigation that are considered “superior strategies” to deterrence. The proposed strategies influence the terrorism by “decreasing the utility of terrorism to terrorists” and by “attempting to increase the opportunity cost”. The strategies are polycentricity, diffusing the attention of the media and offering positive incentives. Polycentricity focuses on creating more than one core/center for the system. In a case where one part of the system is attacked or negatively impacted, the other parts can take control. This is expected to reduce the vulnerability of the overall system and thus decrease the utility of terrorism to terrorists. The concept of diffusing the media suggested by the authors is by the government curtailing credit given to terrorist groups for acts of terrorism in the media. This is expected to reduce the utility of terrorism to the terrorists since gaining public recognition is a major factor for terrorist. However, this strategy is a double-edged sword since it creates media confusion in the community. Individuals will not be completely informed on whether there are more than one existing terrorist groups on attacks or the extent to which their safety is threatened, which can lead to further distrust for the media and government. The last strategy involves giving positive incentives to the terrorists as a mitigation method by “reintegrating terrorists and providing access to the political process, and welcoming repentants” (BBC News, 2020). Although this strategy might be welcoming to the terrorists since it includes access to political processes (which is usually the aim of the attacks) and pardon to “repentants”, a number of questions arise and must be considered. How does the government decide on the extent of political access to provide to the terrorists? How does the government ensure that “repentants” truly have repented? Do the “repentants” face the judicial system? Will the community accept “repentants” back after the damage they have already caused? These questions are extremely critical to the success of this strategy to increase the opportunity cost.

To provide a closer look at the driving forces behind the Boko Haram Insurgency, Alozieuwa (2012) examined various theories, including the Relational/Vengeance Theory, Human Needs/Socio-Economic Perspective, Political Feud Perspective, Islamic Theocratic State Theory and the Conspiracy Theory. After careful examination, it was concluded that originally the Boko Haram attacks were solely driven by religion but evolved to being influenced heavily by politics. In an attempt to develop thorough understanding of the political implication of Boko Haram attacks on Nigeria, Chukwurah et al. (2015) conducted a study. It was gathered that there was an increment of Boko Haram activities as a president from Southeastern Nigerian was inaugurated in 2010. Six sectors were examined during this research. They are the tourism, transportation, infrastructural, commercial, core service and agricultural sectors. The transportation agencies experienced a drastic decline in passengers travelling to the North Nigeria with the uprising of Boko Haram attacks. Tourism from the North, which produced approximately 80 billion naira yearly (the highest in Nigeria), experienced a standstill and then a decline. Infrastructure have been severely damaged, and both foreign and local contractors assigned to be engaged in the rehabilitation of these infrastructure have fled those states due to concern for their safety. The commercial sector, which comprised of corner shops, indigenous markets and small-to-medium businesses, experienced approximately a 73.7% decline and people live on their savings or migrate. The core services such as schools, hospitals, hotels and parks are moving their businesses to other parts of the country. The agricultural sectors are affected since farm lands are no longer being tended to and farmers are concerned for their safety.

Access to improved education, employment, infrastructures such as healthcare facilities and thus a better standard of living might mitigate terrorism in the Northern parts of Nigeria since the poverty and illiteracy levels are relatively high in those regions. However, the members of this terrorist group are in fact exposed to Islamic education, which when misinterpreted could influence them to develop a sense of misguided purpose to facilitate the forceful spread of the religion by any necessary means. Adesoji (2010) found that exposure to Western education was not the solution since “For them, it is a passion, a belief system. I do not believe that exposure to formal western education is the key to mitigate these terrorists”. On the other hand, Adetoro (2012) suggested that thorough state reforms focusing on both infrastructural and structural and considerable poverty relief programs will curtail Boko Haram attacks and other ethno-religious crises in Nigeria. This was concluded after a study was conducted considering poverty and political alienation as major indicators for the Boko Haram Insurgency. Although poverty might be a major factor to consider, the Boko Haram terrorist group might be solely driven by greed and personal vendetta. These are not strange grounds in Nigeria especially when the pursuit of power and political motives are involved.

Mitigating acts of terrorism could be facilitated by considering all factors available and necessary. The game-theoretic approach is a mathematical method of analysis used to understand the interactions between two or more players. This approach considers possible actions from each player and then uses payoffs to determine the optimality of the move. Generally, the government should combat terrorists as that serves the society far more than when no action is taken; however, this is not always the case (Daniel, 2005). Ochoche (2013) performed game theoretic studies using four models: 1, 2A, 2B and 3. In the first model, which is zero-sum, the equilibrium strategy for both the terrorist group and the government were to terrorize and combat, respectively. The second model considered a scenario where the home zone states increased the cost of terrorism and initiated cease fire incentives. The equilibrium strategy remains the same. Model 2B considered increasing the cost of terrorism (a decrease in the payoff). This resulted in cease fire and combat being the dominant strategies for the terrorist group and government, respectively. Ochoche’s fourth model analyzed the condition that the terrorist group has an understanding with the home zone state and moves operations to other states excluding that home zone state from attacks. The dominant strategies will be to terrorize and compromise for the terrorist group and state government, respectively. This approach endangers another state, and does not solve the terrorism problem for the nation or the home zone state. Also, there are no guarantees that the terrorists will continue to comply with the agreement. However, it is unclear how an agreement is reached. The proposed counter-terrorism strategies include increasing the cost of terrorism, initiating cease fire and hindering alliance between state government and terrorist groups. This paper only considers simultaneous games between the government and the terrorist and the payoffs are determined with notional data.

THEORETICAL DEVELOPMENT/MODEL

Game theory is a mathematical method that studies the strategically interdependent behavior between players (Barron, 2013). Interdependence refers to the fact that an action from one player affects the other and vice versa. For this paper, three players are considered: the government, community and the terrorists. The payoffs will be determined based on literature review and analysis of their interests.

The government has two options of combating and engaging the community. The terrorist group has the options of attacking and not attacking. The community takes the role of supporting either the government or the terrorist group. The game tree is presented in Figure 1.

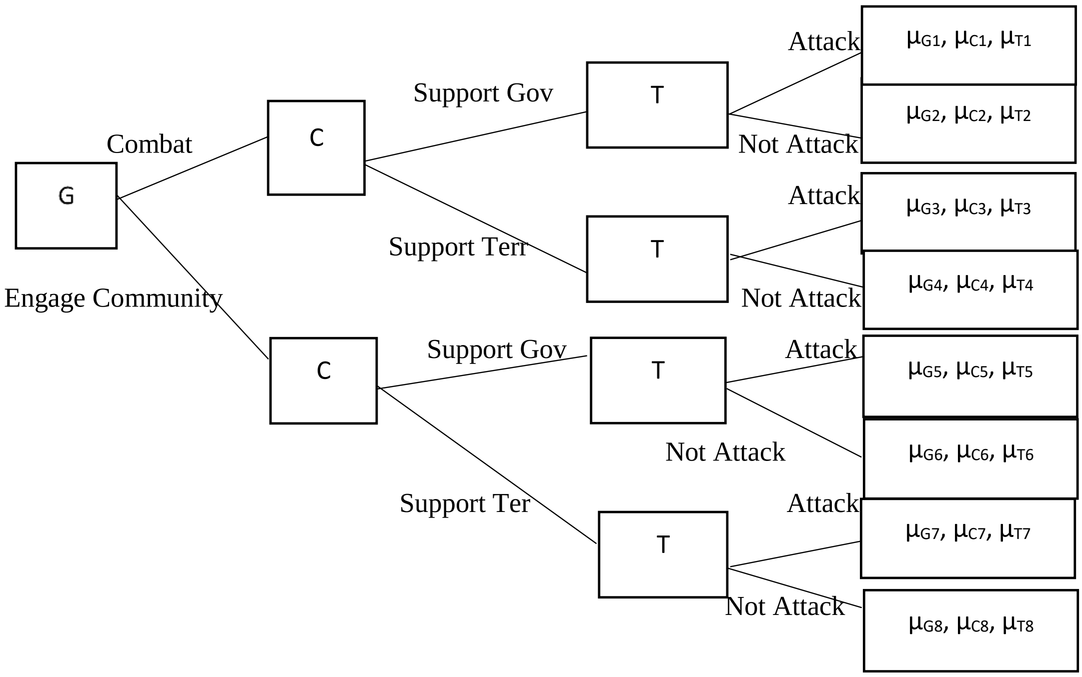


Figure 1. Game tree of the three players: the government, terrorist and community with their payoffs.

Notation

The n-tuple (player) game is mathematically presented as

$$G = (N, S, \mu)$$

where $N = \{1,2,3,\ldots,n\}$ is the set of players and $S = \{S_1, S_2, S_3\ldots, S_n\}$ and $\mu = \{\mu_1, \mu_2, \mu_3,\ldots,\mu_n\}$ are the strategy and payoff sets, respectively.

Every player can perform an action of A_i

Assume $m_i = (a_i^0, a_i^1, a_i^2,\ldots, a_i^{n-1})$ *to be the sequential actions taken by player* $i \in N$.

The zero-sum game in a tactical game is one that for all

$s = \{s_1, s_2, s_3,\ldots,s_n\} \in S$,

$$\mu_1(s) + \mu_2(s) + \mu_3(s) + \ldots + \mu_n(s) = 0$$

All players play to maximize their payoffs. Two strategies s_i and s_i^0 are given so that in any strategy combination, the result from s_i is greater than that from s_i^0. In a set of $s_1, s_2, s_3,\ldots,s_n$, if

$$\mu_i(s_1, s_2, s_3, s_i\ldots, s_n) \geq \mu_i(s_1, s_2, s_3, s_i^0 \ldots, s_n)$$

$s_i \in S$, $where\ s_i$ represents player i's strategies
Nash equilibrium is reached at $\boldsymbol{s} = \{s_1, s_2, s_3, \ldots, s_n\}$

Justification of the Government's Choices

Government can choose to either heighten security or engage the community. If the government chooses to heighten security, then they must consider defense strategy against the terrorists, cost of the defense, and the impact factor. Generally, the government aspires to target the terrorists while minimizing damage to the community including infrastructure damages, casualty, societal tie disruptions, and so on.

According to the United Nations, the government heightens security by securing both domestic and international borders, constricting financial regulations, increasing the involvement of police authorities, improving the criminal justice system, and establishing legal alliance with other countries with terrorist threats to help convict terrorists in their courts (Smith, 2020). By defending, the government may monitor websites and online contents, to take down hate speech and suspicious activities. The Dutch government has implemented numerous actions to strengthen its defense against terrorism. The Royal Netherlands Air Force surveys the Dutch airspace on a 24 hours basis. Also, the police closely monitor people, who they suspect to be terrorist threats to the society (Government of the Netherlands, 2016-2020). The 'Counterterrorism Alert System' established by the Dutch government notifies the government and primary sectors about terrorist threats (2016-2020). These sectors including drinking water, and energy companies.

Other impactful strategies by the government to form defense against terrorism can be exemplified through the North Atlantic Treaty Organization (NATO). NATO narrows efforts towards terrorism mitigation through improving awareness of terrorist threats, setting up capabilities to prepare and adequately respond to threats, and forming alliances/partnerships with neighboring countries and other international actors (NATO, 2019).

According to Zycher (2003), annual costs of deaths and injuries from moderate, severe, and nuclear cases in the United States were estimated at $11 billion, $183 billion, and $465 billion, respectively, before the September 11 attacks. These values are based on the estimates that one life is worth $4 million, and one injured individual is worth $40,000. However, after the September 11 attacks, government spending for reconstructing, humanitarian activities, defense, economic aid, and domestic security functions have been estimated at $95 billion.

The government can engage the community by establishing community enrichment programs, and protecting and rewarding informants. The Office for Victims of Crime (OVC) in the U.S. Department of Justice provides 24/7 crisis counselling services in English and Spanish to victims of terrorism and mass violence through the "Terrorism and Special Jurisdictions Program" (Office for Victims of Crime, 2020). Also, OVC provides victim compensation to the affected states for the welfare of the affected within the state. Other organizations that provide support are "The Dougy Center", "National Organization of Parents of Murdered Children", and "VictimConnect" (Office for Victims of Crime, 2020). The United States' government through the Federal Bureau of Investigation (FBI) established the FBI Counterterrorism Center in 1996 (Watson, 2002) for combating terrorist activities on both domestic and international levels.

The Nigerian government has made efforts to engage the community through providing incentives to the community for disengage in violent acts. Recently, the Zamfara State Governor, Bello Matawalle, has offered two cows to the indigenous people for every AK-47 or weapon returned to the government (BBC News, 2020). Zamfara is a state in northwestern Nigeria with approximately 67.5% people living in poverty (BBC News, 2020). A lucrative business in Zamfara State is farming, especially animal herding, and cows are highly valued by the average herdsman. Also, over 8,000 people have been killed through crimes related to terrorism, theft, and inter-ethnic tension in Zamfara and its neighboring states, which makes this incentive not only relevant but necessary. Another way that the government can engage the community is through organizations such as the U.S. Intelligence Community (IC) Civilian Joint Duty Program, which offers civilians with professional opportunities to strengthen collaboration with the government and community. Also, the program provides the participating civilians with enhanced career prospects through trainings and exposure to the processes involved in intelligence (Office of the Director of National Intelligence, 2020).

Exploring Different Options for the Community

The community can choose either to support the government or support the terrorist. The community can support the government by reporting suspicious activity to the police authorities, become involved in campaigns organized by the government to influence the children and youth, and refrain from voluntary recruitment by the terrorist. In 2007, the U.S. government established the Nationwide Suspicious Activity Reporting (SAR) Initiative (NSI) which is a collective effort by the Federal Bureau of Investigation, Department of Homeland Security, state, local, tribal, and territorial law enforcement partners (Joint Counterterrorism Assessment Team, 2020). This initiative informs on how to safely report suspicious activities to the appropriate authorities.

Payoff Functions

To derive a better understanding of the players, their decisions, and outcomes and thus payoff functions, their definitions and other notations are outlined in Table 1.

Table 1. Players, decisions, payoff and their definitions including other notations

Players	*Definition*
G	Government
C	Community
T	Terrorist
Decision Variables	*Definition*
{H, EC}	Government heightens security or engages community
{SG, ST}	Community supports government or terrorist
{A, NA}	Terrorist attacks or not
Government's Payoffs	
Parameters	*Definition*
L_G	Impact factor to government: cost of damages to the government after defense
I_G	Impact to government for engaging the community when the terrorist attacks
J_G	Loss to government when government heightens security and community supports terrorist
H_G	Loss to government when government engages community and community supports terrorist

V_G	Benefit to government when government engages community and community supports government
C_D	Cost of defense to the government
C_E	Cost of engaging the community to the government
Community's Payoffs	
Parameters	*Definition*
S_E	Impact to community when government engages community and community supports government
S_W	Impact to community when government engages community and community supports terrorist
S_G	Impact to community for supporting government when government heightens security
S_T	Impact to community for supporting the terrorist when government heightens security
I_C	Impact to community for supporting the government when terrorist attacks
T_C	Impact to community for supporting the terrorist when terrorist attacks
X_C	Benefit to community for supporting government when terrorist does not attack
P_C	Loss to community for supporting terrorist when terrorist does not attack
Terrorist's Payoffs	
Parameters	*Definition*
L_T	Impact factor to terrorist: cost of damages done by the terrorist after attack
O_T	Impact to terrorist when government engages community when the terrorist attacks
K_T	Impact on terrorist when community supports government and terrorist attacks
B_T	Benefit to terrorist when community supports terrorist and terrorist attacks
MC	Media coverage for terrorist attacks
C_A	Cost of attack to the terrorist
Other Notations	
{Y, N}	Possible Equilibrium or Not a Possible Equilibrium, respectively
{μ}	Assigned payoff

Considering the Two Player Game between Subgames

We first consider two-player sequential and simultaneous games between the government and terrorist, the government and community, and the community and terrorist, respectively.

Government and Terrorist

Figure 2 shows the sequential game between the government and terrorist. In this sequential game, the government makes the first move, and the terrorist is aware of the government's move before their move is made. When the government heightens security (H), the payoff is $-\mathrm{L_G} - \mathrm{C_D}$ or $-\mathrm{C_D}$ depending on if the terrorist attacks or not, respectively. When the government

engages community (EC), the payoff is
$-\mathrm{I_G} - \mathrm{C_E}$ or $-\mathrm{C_E}$ depending on if the terrorist attacks or not, respectively. On the other hand, the terrorist payoff is $L_T + MC - C_A$ or 0 depending on if they attack or not when the government heightens security. Similarly, the terrorist payoff is $L_T + O_T + MC - C_A$ or 0 depending on if they attack or not when the government engages community.

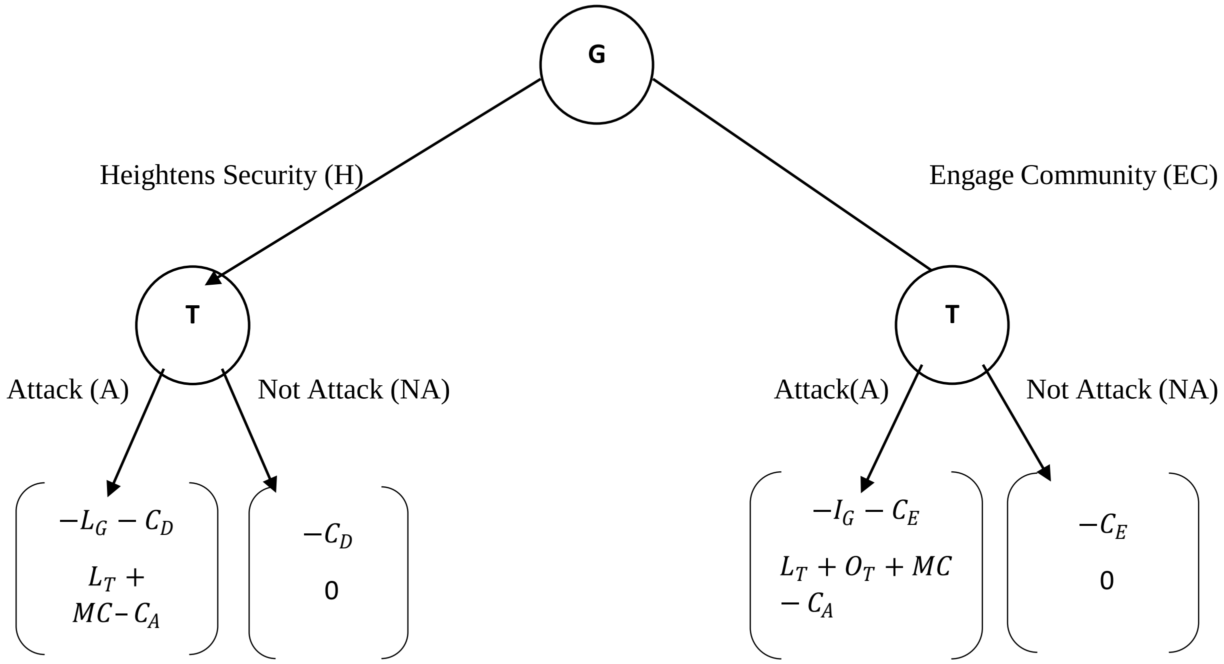


Figure 2. Game tree of the sequential subgame between the government (G) and terrorist (T).

Table 2. Possible equilibria for subgame between government and terrorist

Cases	Strategies	Payoffs	Conditions
1	[H, A]	$[-L_G - C_D, L_T + MC - C_A]$	$\{C_A < L_T + MC, C_A < L_T + O_T + MC,$ $C_E > L_G + C_D - I_G\}$
2	[H, NA]	$[-C_D, 0]$	$\{L_T + O_T + MC < C_A < L_T + MC,$ $C_E > L_G + C_D\}$ *or* $\{L_T + MC < C_A < L_T + O_T + MC,$ $C_E > C_D - I_G\}$ *or* $\{C_A > max\{L_T + MC, L_T + O_T + MC\},$ $C_E > C_D\}$
3	[EC, A]	$[-I_G - C_E, L_T + O_T + MC - C_A]$	$\{C_A < L_T + MC, C_A < L_T + O_T + MC,$ $C_E < L_G + C_D - I_G\}$
4	[EC, NA]	$[-C_E, 0]$	$\{L_T + O_T + MC < C_A < L_T + MC,$ $C_E < L_G + C_D\}$ *or* $\{L_T + MC < C_A < L_T + O_T + MC,$ $C_E < IG - C_D\}$ *or* $\{C_A > max\{L_T + MC, L_T + O_T + MC,$ $C_E < C_D\}$

Solution

In the sequential game, from solving the subgame between the government and terrorist, all four cases present feasible equilibria provided the conditions are met (as shown in Table 2). They are:

1. When the cost of attack is low ($C_A < min\,\{L_T + MC,\, L_T + O_T + MC\}$) when the community supports either the government or the terrorist, and the cost of engaging the community to the government are high ($C_E > L_G + C_D - I_G$), the government will heighten security and terrorist will attack.
2. When the cost of attack is medium ($L_T + O_T + MC < C_A < L_T + MC\ if\ O_T < 0;\ or\ L_T + O_T + MC < C_A < L_T + MC\ if O_T \geq 0$) and the cost of engaging the community to the government are high ($C_E > L_G + C_D$), or when the cost of attack is medium ($L_T + MC < C_A < L_T +\ O_T + MC\ \ if O_T \geq 0$) and the cost of engaging the community to the government is relatively high ($C_E > C_D - I_G$), or when the cost of attack is high ($C_A > max\{L_T + MC, L_T + O_T + MC\}$) and the cost of engaging the community to the government is high ($C_E > C_D$), the government will heighten security and terrorist will not attack.
3. When the cost of attack is low ($C_A < \max\{L_T + MC,\, L_T + O_T + MC\}$), the community supports either the government or the terrorist and the cost of engaging the community to the government is low ($C_E < L_G + C_D - I_G$), the government will engage the community and terrorist will attack.
4. When the cost of attack is medium ($L_T + O_T + MC < C_A < L_T + MC\ if\ O_T < 0$) and the cost of engaging the community to the government are low ($C_E < L_G + C_D$), or when the cost of attack is medium ($L_T + MC < C_A < L_T +\ O_T + MC\ if O_T \geq 0$) and the cost of engaging the community to the government are relatively low ($C_E < L_G + C_D$), or when the cost of attack is high ($C_A > max\{L_T + MC, L_T + O_T + MC\}$), and the cost of engaging the community to the government is low ($C_E < C_D$), the government will engage the community and terrorist will not attack.

However, considering that this subgame is a sequential game where the terrorist is aware of the government's moves before their move is selected, the most desirable outcomes for the government will be one that minimizes the impact of the damage to the community by the terrorist, and defends the community and nation as a whole. Hence, the most desirable equilibrium for the government will be [H, NA] and [EC, NA] where the terrorist does not attack provided the conditions in Table 2 for these cases are met. [EC, A] will not be a desirable equilibrium for the government since when the government engages the community, the terrorist will prefer to attack to get a higher payoff than otherwise which is not desirable to the government under the given conditions. [H, A] will not be a desirable equilibrium for the government since when the government heightens security, the terrorist will prefer to attack to get a higher payoff than otherwise which is not desirable to the government under the given conditions.

For the terrorist, the most desirable equilibrium is the one where the most damage is achieved. Hence, the design equilibria for the terrorist will be [H, A] and [EC, A], where the terrorist attacks provided the conditions in Table 2 for these cases are met. [EC, NA] will not be a desirable equilibrium for the terrorist since if the government chooses to engage the community, the terrorist will prefer not to attack to receive a higher payoff than otherwise under the given conditions. [H, NA] will not be a desirable equilibrium for the terrorist since if the government chooses to heighten security, the terrorist will prefer not to attack to get a higher payoff than otherwise under the given conditions.

Next, considering a simultaneous game where neither player knows the moves of the other player, both players' moves are made at the same time solely based on the best outcome for the individual player regardless of the other player's move as shown in Table 3. To determine the conditions for equilibrium, we use best response analysis to find the equilibrium. First, we will evaluate the player's payoffs from the government's perspective, then we will evaluate the payoffs from the terrorist's perspective.

Table 3. Simultaneous subgame between the government (G) and the terrorist (T).

	T	
G	Attack (A)	Not Attack (NA)
Heighten Security (H)	$[-L_G - C_D, L_T + MC - C_A]$	$[-C_D, 0]$
Engage Community (EC)	$[-I_G - C_E, L_T + O_T + MC - C_A]$	$[-C_E, 0]$

Table 4. Possible equilibria for simultaneous subgame between government and terrorist

Cases	**Strategies**	**Payoffs**	**Conditions**
1	[H, A]	$[-L_G - C_D, L_T + MC - C_A]$	$\{C_A < L_T + MC, C_A < L_T + O_T + MC, C_E > L_G + C_D - I_G\}$
2	[H, NA]	$[-C_D, 0]$	$\{C_A > max\{L_T + MC, L_T + O_T + MC\}, C_E > C_D\}$
3	[EC, A]	$[-I_G - C_E, L_T + O_T + MC - C_A]$	$\{C_A < L_T + MC, C_A < L_T + O_T + MC, C_E < L_G + C_D - I_G\}$
4	[EC, NA]	$[-C_E, 0]$	$\{C_A > max\{L_T + MC, L_T + O_T + MC, C_E < C_D\}$

Solution

In the simultaneous game, from solving the subgame between the government and terrorist, all four cases present feasible equilibrium provided the conditions are met as shown in Table 4. They are:

1. When the cost of attack is low ($C_A < \min\{L_T + MC, L_T + O_T + MC\}$), and the cost of engaging the community to the government are high ($C_E > L_G + C_D - I_G$), the government will heighten security and terrorist will attack.
2. When the cost of attack is high ($C_A > \max\{L_T + MC, L_T + O_T + MC\}$) and the cost of engaging the community to the government are high ($C_E > C_D$), the government will heighten security and terrorist will not attack.
3. When the cost of attack is low ($C_A < \min\{L_T + MC, L_T + O_T + MC\}$) when the community supports either the government or the terrorist, and the cost of engaging the community to the government is low ($C_E < L_G + C_D - I_G$), the government will engage the community and terrorist will attack.
4. When the cost of attack is high ($C_A > max\{L_T + MC, L_T + O_T + MC\}$) and the cost of engaging the community to the government are low ($C_E < C_D$), the government will engage the community and terrorist will not attack.

Considering that this subgame is a simultaneous game, where the terrorist is not aware of the government's move before their move is selected, and both players select their moves at the same time. Similar to the sequential game, the most desirable outcomes for the government will be the one that minimizes the impact of the damage to the community by the terrorist, and defends the nation as a whole. Hence, the most desirable equilibrium for the government will be [H, NA] and [EC, NA] where the terrorist does not attack provided the conditions in Table 4 for these cases are met. [EC, A] will not be a desirable equilibrium for the government since when the government engages the community, the terrorist will prefer to attack to get a higher payoff than otherwise which is not desirable to the government under the given conditions. [H, A] will not be a desirable equilibrium for the government since when the government heightens security, the terrorist will prefer to attack to get a higher payoff than otherwise which is not desirable to the government under the given conditions.

For the terrorist, the most desirable equilibrium is the one where the most damage is achieved. Hence, the equilibrium for the terrorist will be [H, A] and [EC, A], where the terrorist attacks provided the conditions in Table 4 for these cases are met. [EC, NA] will not be a desirable equilibrium for the terrorist since if the government chooses to engage the community, the terrorist will prefer not to attack to receive a higher payoff than otherwise under the given conditions. [H, NA] will not be a desirable equilibrium for the terrorist since if the government chooses to heighten security, the terrorist will prefer not to attack to get a higher payoff than otherwise under the given conditions.

Comparing the Possible Equilibria for the Sequential and Simultaneous Subgame between the Government and Terrorist

Comparing the payoffs for both the sequential and simultaneous subgames between the government and terrorist, the conditions for possible equilibrium for cases 1 and 3 (the most desirable outcomes for the terrorists) are the same. However, for cases 2 and 4 (the most desirable outcomes for the government), there are more conditions in the sequential game than in the simulations game. This indicates that the government is more likely to reach their desired outcome in a sequential game than in a simultaneous game. Hence, we recommend that the government announce their strategy to the terrorists rather than keep their strategy unknown to the terrorist to better reach their desired outcome.

In Figure 3a, when $O_T \geq 0$, the cost of attack for the terrorist is low, and the cost of community engagement on the government when the terrorist attacks is high, and might become detrimental to the government if engaging the community is not successful.

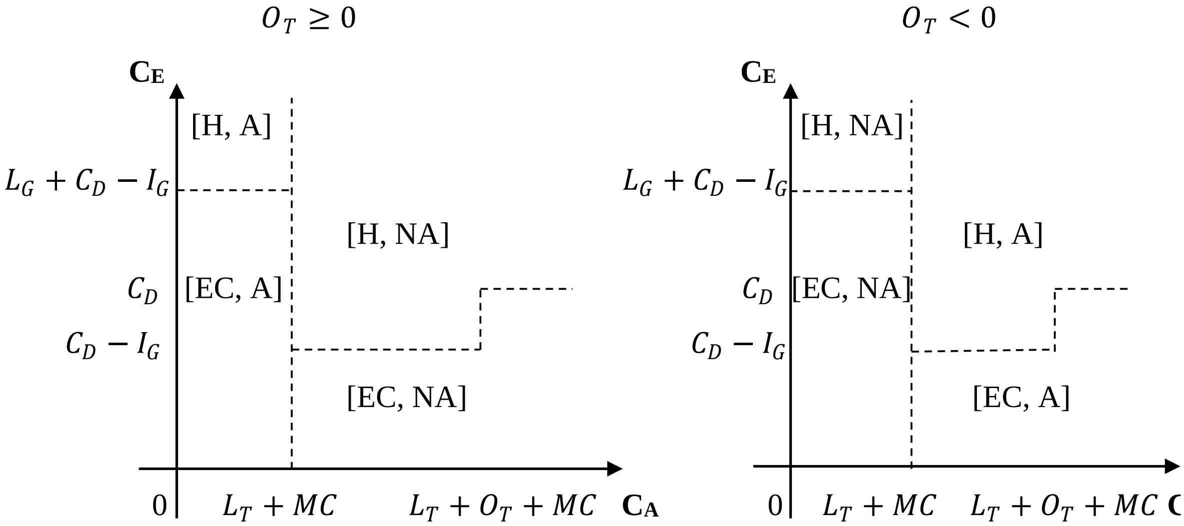


Figure 3. Four possible equilibria (as specified in Table 2) for subgame between government and terrorist as a function of C_A and C_E. The strategies in the brackets are for the government and terrorist respectively.

In Figure 3b, when $O_T < 0$, the cost of attack for the terrorist is high, and the cost of community engagement for the government when the terrorist attacks is relatively low, the government has a higher chance of success with community engagement, while the terrorist will likely not attack when cost of attack is high.

Government and Community

Figure 4 shows the sequential game between the government and community. In this sequential game, the government makes the first move, and the community is aware of the government's move before their move is made. When the government heightens security (H), the payoff is $-C_D$ or $-J_G - C_D$ depending on if the community supports government or supports terrorist, respectively. When the government engages community (EC), the payoff is $V_G - C_E$ or $-H_G - C_E$ depending on if the community supports government or supports terrorist, respectively. On the other hand, the community payoff is S_G or S_T depending on if they support government or terrorist, respectively, when the government heightens security. Similarly, the terrorist payoff is S_E or S_W depending on if they support government or terrorist, respectively, when the government engages community.

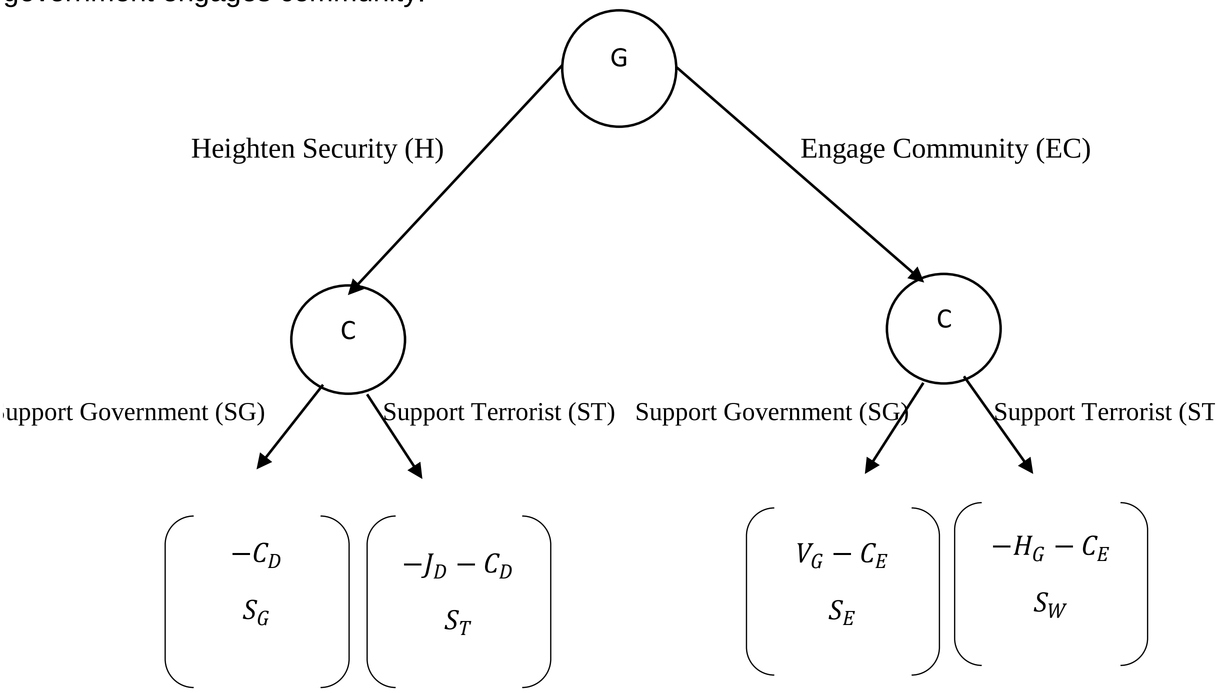


Figure 4. Game tree showing the sequential subgame between the government (G) and the community (C).

Table 4. Possible equilibria for sequential subgame between community and government

Cases	Strategies	Payoffs	Conditions
1	[H, SG]	$[-C_D, S_G]$	$\{S_G \geq S_T,\ S_E \geq S_W,\ C_D \leq -V_G + C_E\}$ or $\{S_G \geq S_T, S_E < S_W, C_D \leq H_G + C_E\}$
2	[H, ST]	$[-J_G - C_D, S_T]$	$\{S_G < S_T, S_E \geq S_W, C_D \leq H_G + C_E - J_G\}$ or $\{S_G < S_T, S_E < S_W, C_D \leq -V_G + C_E - J_G\}$
3	[EC, SG]	$[V_G - C_E, S_E]$	$\{S_G \geq S_T,\ S_E \geq S_W,\ C_D > -V_G + C_E\}$, or $\{S_G < S_T, S_E \geq S_W, C_E < V_G + C_D + J_G\}$
4	[EC, ST]	$[-H_G - C_E, S_W]$	$\{S_G < S_T, S_E < S_W, C_D > H_G + C_E - J_G\}$ or $\{S_G > S_T, S_E < S_W, C_D > H_G + C_E\}$

Sequential Game Solution

From solving the sequential subgame between the government and community, all four cases present feasible equilibria provided the conditions are met as shown in Table 4. They are:

1. When the impact on community of supporting government when government heightens security is higher than that of supporting the terrorist ($S_G \geq S_T$), the impact on community for supporting the government when government engages the community is higher than that of supporting the terrorist ($S_E \geq S_W$), and when the cost of defense is low, and the cost of engaging the community to the government is high ($C_D \leq \ min\{-V_G + \ C_E, H_G + C_E\}$), the government heightens security and community supports the government.
2. When the impact on community for supporting government when government heightens security is low ($S_G < S_T$), the impact on community for supporting the terrorist when government heightens security is high ($S_E \geq S_W$), and when the cost of defense is low and the cost of engaging the community to the government is high ($C_D \leq \ min\{-H_G + C_E - J_G, -V_G + C_E - J_G\}$), the government will heighten security and community will support the terrorist.
3. When the impact on community for supporting government when government heightens security is high ($S_G \geq S_T$), the impact on community for supporting the terrorist when government heightens security is high ($S_E \geq S_W$), and when the cost of defense is high and the cost of engaging the community to the government is low ($C_D \geq \ -V_G + C_E$), the government will engage the community and the community will support the government.
4. When the impact on community for supporting government when government heightens security is low ($S_G < S_T$), the impact on community for supporting the terrorist when government heightens security is low ($S_E < S_W$), and when the cost of defense is high, and the cost of engaging the community to the government is low ($C_D \geq max\{-H_G + C_E - J_G, -H_G + C_D\}$), the government will engage the community and the community will support the terrorist.

Considering that this subgame is a sequential game, where the community is aware of the government's move before their own move is selected, the most desirable outcomes for the government will be the one, where the community supports the government regardless of the government's move. Hence, the most desirable equilibrium for the government will be where the community supports the government [H, SG] and [EC, SG]. Additionally, [H, SG] and [EC, SG] will be desirable equilibria provided the conditions in Table 6 for these cases are met.
For the community, the most desirable equilibrium will be the one where the least damage is achieved. Hence, the equilibrium for the community will be [H, SG], [H, ST], [EC, SG], and [EG, ST] where the community supports either the terrorist or the government depending on how the decision of either player affects the community provided the conditions in Table 6 for these cases are met.

Then, we consider a simultaneous game as shown in Table 5, where neither player knows the move of the other player, both players' moves are made simultaneously solely based on the best outcome for the individual player regardless of the other player's move. We use best response analysis to find the equilibrium. First, we evaluate the player's payoffs from the government's perspective. Then, we evaluate the payoffs from the community's perspective.

Table 5. Payoffs of the simultaneous subgame between the government (G) and community (C)

	C	
G	Support Government (SG)	Support Terrorist (ST)
Heighten Security (H)	$[-C_D, S_G]$	$[-J_G - C_D, S_T]$
Engage Community (EC)	$[V_G - C_E, S_E]$	$[-H_G - C_E, S_W]$

Table 6. Possible equilibria for simultaneous subgame between government and community

Cases	Strategies	Payoffs	Conditions
1	[H, SG]	$[-C_D, S_G]$	$\{C_D \leq -V_G + C_E, S_G \geq S_T, S_E \geq S_W\}$
2	[H, ST]	$[-J_G - C_D, S_T]$	$\{C_D \leq H_G + C_E - J_G, S_G < S_T, S_E < S_W\}$
3	[EC, SG]	$[V_G - C_E, S_E]$	$\{C_D > -V_G + C_E, S_G \geq S_T, S_E \geq S_W\}$
4	[EC, ST]	$[-H_G - C_E, S_W]$	$\{C_D > H_G + C_E - J_G, S_G < S_T, S_E < S_W\}$

Simultaneous Game Solution

In the simultaneous game, from solving the subgame between the government and community, all four cases present feasible equilibrium provided the conditions are met (as shown in Table 6 and Figure 5).

1. When the impact on community for supporting government when government heightens security is high $(S_G \geq S_T)$, the impact on community for supporting the government when government engages the community is high $(S_E \geq S_W)$, and when the cost of defense is low, the cost of engaging the community to the government is high $(C_D < -V_G + C_E)$, the government will heighten security and community will support the government.
2. When the impact on community for supporting government when government heightens security is low $(S_G < S_T)$, the impact on community for supporting the terrorist when government engages the community is high $(S_E < S_W)$, the cost of defense is low and the cost of engaging the community to the government is high $(C_D < H_G + C_E - J_G)$, the government will heighten security and community will support the terrorist.
3. When the impact on community for supporting government when government heightens security is high $(S_G \geq S_T)$, the impact on community for supporting the terrorist when government engages the community is low $(S_E \geq S_W)$, the cost of defense is high and the cost of engaging the community to the government is low $(C_D \geq -V_G + C_E)$, the government will engage the community, and the community will support the government.
4. When the impact on community for supporting government when government heightens security is low $(S_G < S_T)$, the impact on community for supporting the terrorist when government engages the community is high $(S_E < S_W)$, the cost of defense is low and the cost of engaging the community to the government is low $(C_D \geq -H_G + C_E - J_G)$, the government will engage the community and the community will support the terrorist.

Considering that this subgame is a simultaneous game, where the community is not aware of the government's moves before their move is selected, and both players select their moves at the same time. Similar to the sequential game, the most desirable outcomes for the government will be [H, SG], and [EC, SG], where the government takes an action to either heighten security or engage community provided the conditions in Table 6 for these cases are met. [EC, ST] is not a desirable equilibrium for the government since when the government engages the community, the community prefers to support the terrorist to receive a higher payoff than otherwise which is not desirable to the government under the given conditions. [H, ST] is not a desirable equilibrium for the government since when the government heightens security, the community prefers to support the terrorist to receive a higher payoff than otherwise under the given conditions.

For the community, the most desirable equilibrium will be the one, where the least damage is achieved. Hence, the desirable equilibrium for the community will be [H, SG], [EC, SG], and [EG, ST] where the community supports either the terrorist or the provided the conditions in Table 6 for these cases are met. [H, ST] is not a desirable equilibrium for the community since if the government chooses to heighten security, the community will be at a loss if they support the terrorist especially if the terrorist does not attack. Hence, the community will prefer to support the government to receive a higher payoff than otherwise under the given conditions.

Comparing the Possible Equilibria for the Sequential and Simultaneous Subgame between the Government and Terrorist

While comparing the payoffs for the sequential and simultaneous subgames between the government and community, we notice that the conditions for possible equilibrium for all cases exists under one set of conditions in the simultaneous game. However, there are two sets of conditions to achieve the given payoffs for all cases in the sequential game. This indicates that the both players, are more likely to reach their desired outcome in the sequential game than in the simultaneous game. Therefore, we recommend that the government announce their strategy to the community rather than keep their strategy unknown to the community to better reach their desired outcome vice versa.

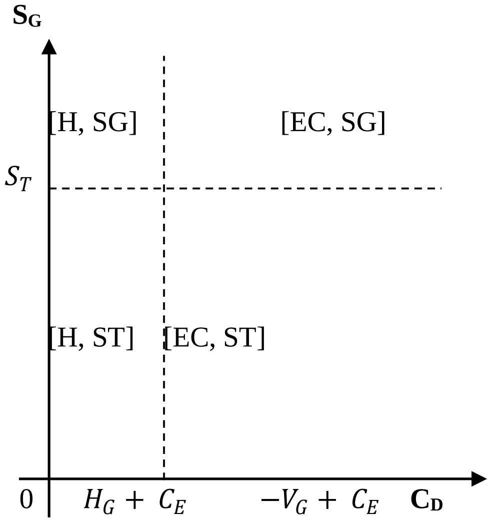


Figure 5. Four possible equilibria (as specified in Table 6) for subgame between government and terrorist as a function of C_D and S_G

CONCLUSION

In this research, game theory was used as a means to analyze the interactions between the government, terrorists and the community with regards to terrorism mitigation. Two subgames between the government and terrorist, and between the government and community, respectively were developed and solved. For both subgames, sequential and simultaneous games are compared. We identified possible equilibria and their leading conditions for both subgames in both orders of moves. We also analyzed the most desirable equilibria for all subgames. We compared the sequential and simultaneous games and confirmed the first mover advantage in both the sequential game between the government and terrorist (when the government moves first) and that between the government and community (when the government moves first) supporting previous research (Zhuang and Bier 2007). While studying the two-player sequential and simultaneous subgames games, the results show that all three players have more conditions to reach possible equilibrium in the sequential game rather than in the simultaneous game. This leads to the first conclusion that the government, make their strategy publicly available to the community, and terrorist in order to receive their desired outcomes rather than keeping it undisclosed.

This topic is relevant for mitigating terrorism since terrorism is a growing world challenge, which endangers innocent citizens, increases crime rates, and deteriorates the economy of concerned nations. The community being the most vulnerable sub-unit of the nation could experience high displacement into already relatively saturated communities due to terrorism in some countries. The rise of lone wolf terrorist acts suggests the importance of engaging the community so that they are likely to support the terrorist or becoming a terrorist. Therefore, it is important to find efficient ways to stop the community from supporting terrorists out of necessity to survive.

While it would be extremely difficult to (approximately) estimate the parametric costs and benefits, it could be instrumental to provide more viable predictions regarding the decisions of the three key stakeholders. Among all those uncertain parameters, costs might be relatively easier to estimate while benefits such as reduction in terrorist attack loss due to community engagement would be almost infeasible to quantify given very limited data are available. However, the externalities of some community engagement strategies could reduce the crime rates among other benefits might be possible. Another interesting future research question is regarding the generalizability of this modeling framework to particular country settings. The model parameters are likely to differ for various communities/countries and involved terrorist groups. In the future, we could also consider a continuous model where the decision variable of community engagement is continuous and possibly yield insights into the optimal level of community engagement in combating terrorism.

REFERENCES

Adesoji, A. (2010). The Boko Haram Uprising and Islamic Revivalism in Nigeria. *Africa Spectrum*, 100-104.

Adetoro, R. A. (2012, November-December). Boko Haram insurgency in Nigeria as a symptom of poverty and political alienation. *Journal of Humanities and Social Science, 3*(5), 21-26.

Alozieuwa, S. H. (2012). Contending Theories on Nigeria’s Security Challenge in the Era of Boko Haram Insurgency. *Peace & Conflict Review, 7*(1), 1-8.

Barron, E. N. (2013). *Game Theory : An Introduction* (2 ed., Vol. 1). Hoboken, NJ: John Wiley & Sons.

Boettke, P. J., Lemke J. S., & Palagashvili, L. (2013). Riding in cars with boys: Elinor Ostrom's adventures with the police. *Journal of Institutional Economics 9*(4), 13-01.

Boettke, P. J., Lemke J. S., & Palagashvili, L. (2016) Re-evaluating community policing in a polycentric system. *Journal of Institutional Economics 12*(2): 305-325.

BBC News. (2020, July 17). Nigeria's Zamfara state offers repentant bandits cows for AK-47s . *BBC News - Africa*.

Boudreau, E. P. (2020, March 13). *Supporting One Another in Times of Crisis.* Retrieved from Harvard Graduate School of Education: https://www.gse.harvard.edu/news/uk/20/03/supporting-one-another-times-crisis, September 28, 2021.

Charles L., R. (2002). The Definition of Terrorism. *Analyses of Social Issues and Public Policy*, 9-14.

Chukwurah, D. C. (2015, May). Implication of Boko Haram Terrorism on Northern Nigeria. *Mediterranean Journal of Social Sciences, 6*(3), 371-379.

CNN Library. (2019). *Boko Haram Fast Facts.* CNN World . Retrieved from https://www.cnn.com/2014/06/09/world/boko-haram-fast-facts/index.html December 7, 2019.

Daniel, G. A. (2005, April). Counterterrorism: A Game-Theoretic Analysis. *Journal of Conflict Resolution, 49*(2), 183 - 200.

Dragos, A., P. and P. J. Boettke. (2009). *Challenging institutional analysis and development: The Bloomington school.* Routledge.

Escalante, E. E. (2020). Night Watchers and Terrorists. *Journal of Economic Behavior & Organization 171*: 116-131.

Fleming, J. (2005). 'Working together' : Neighborhood Watch, reassurance policing and the potential of partnerships. *Australian Institute of Criminology*.

Goodman, N. (2018). The Coproduction of Justice. In: *Rethinking Punishment in the Era of Mass Incarceration*, ed. Chris W. Surprenant (Routledge): 49-68.

Harvard Health Publishing. (2020, May 13). *Preventing the spread of the coronavirus .* Retrieved from Harvard Medical School : https://www.health.harvard.edu/diseases-and-conditions/preventing-the-spread-of-the-coronavirus, September 28, 2021.

Institute for Economics & Peace. (2019, November). *Global Terrorism Index 2019: Measuring the Impact of Terrorism.* Retrieved from http://visionofhumanity.org/reports, September 28, 2021.

Laqueur, W. (1987). *The Age of Terrorism .* Boston : Little Brown.

Ochoche, J. M. (2013, November). Modeling Boko Haram: A Game Theoretical Approach. *International Journal of Science and Technology, 2*(11), 777-783.

Office of the Director of National Intelligence. (2020). *Joint Duty .* Retrieved from Office of the Director of National Intelligence: https://www.dni.gov/index.php/who-we-are, September 28, 2021.

Passas, N. (2005). *Informal Value Transfer Systems, Terrorism and Money Laundering.* Boston: U.S. Department of Justice.

Richardson, Harry W., Gordon, Peter, Moore, James Elliott. (2007). *The Economic Costs and Consequences of Terrorism.* Cheltenham, UK: Edward Elgar Publishing.

Shan, X. and Zhuang, J. (2014, March). Modeling Credible Retaliation Threats in Deterring the Smuggling of Nuclear Weapons Using Partial Inspection - A Three-Stage Game. *Decision Analysis, 11*(1), 43-62.

START. (2009-2019). *Global Terrorism Database*. Retrieved from National Consortium for the Study of Terrorism and Responses to Terrorism: https://www.start.umd.edu/gtd/search/Results.aspx?search=IDP+Nigeria&sa.x=10&sa.y=14, September 28, 2021.

Vlad, T. (2016). *Elinor Ostrom: an intellectual biography.* Rowman & Littlefield.

United Nations High Commissioner for Refugees. (2001-2020). *Nigeria Emergency*. Retrieved from UNHCR: https://www.unhcr.org/en-us/nigeria-emergency.html, September 28, 2021.

Watson, D. L. (2002, February 06). *Before the Senate Select Committee on Intelligence .* Retrieved from The Federal Bureau of Investigation: https://archives.fbi.gov/archives/news/testimony/the-terrorist-threat-confronting-the-united-states, September 28, 2021.

Woodhead, J. (1990). The Role of the Community in Crime Prevention - Neighbourhood Watch and the broader Community Interests. *Australian Institute of Criminology - The Police And The Community In The 1990s*, 84-85.

Zhuang, J., & Bier, V. M. (2007). Balancing Terrorism and Natural Disasters—Defensive Strategy with Endogenous Attacker Effort. *Operations Research*, 55(5), 976-991.

DECISION SCIENCES INSTITUTE

The Application of Natural Language Processing as a Tool for Knowledge Extraction for COVID-19.

Ganga Prasad Basyal
Dakota State University
Email: ganga.basyal@trojans.dsu.edu

Piyush Vyas
Dakota State University
Email: piyush.vyas@trojans.dsu.edu

Dave Bishop
Dakota State University
Email: dave.bishop@dsu.edu

ABSTRACT

During the span of the last two-year numerous studies has been published and the social media is overwhelmed with the news related to Covid-19. Goal of our research is to extract the knowledge from this huge pile of data. To extract the knowledge, we utilized the Twitter data containing 1,75,000 Tweets. Exploratory Data Analysis (EDA) and prominent text mining and analysis techniques are applied to generate the results. We applied topic mining technique to generate results in form of highly relevant topics, converting raw information to knowledge. Results represented as knowledge dimensions will contribute for making informed decision.

KEYWORDS: Covid-19, Social media analytics, Knowledge management, NLP, decision support systems.

INTRODUCTION

At the end of 2019, a novel virus had been detected through the sample swap from one of the patients in China, confirming a new type of SARS virus named COVs (Ge et al, 2020) later named as Covid-19 or Coronavirus. There are many speculations about the origin of Covid-19 in the research community. But many research studies unanimously pointed towards China and to be more specific, a Wuhan seafood market in China, as the origin of Covid-19 (Mehta et al, 2020), (Guo et al, 2020).

Given the severity of the pandemic, it is imperative to gain understanding of precautions related to Covid-19. The medium of social media is one of the biggest sources of public opinion and has been the center of information communication, along with this, the amount of mis-information and its spread in communities has also increased significantly (Shahi et al, 2021). In order to stay informed many studies have been conducted recently using social media data. Facebook, Twitter, and Youtube are some of the top platforms participating in information diffusion (Cao et al, 2020; Cinelli et al, 2020). However, many of these studies have issues regarding their scope of research along the lines of the geography covered, key words for sourcing data (Pokharel, 2020), and platform narrowness (Cao et al, 2020). These issues have result in low sample data the representing entire community. Second, the studies conducted have used text mining methods partially (Giachanou and Crestani, 2016; Kumar and Sebastian, 2012). To overcome these limitations, a comprehensive study is required which includes data from scattered geographic locations and implements complete text mining framework.

Studies have been conducted in past for managing information and its application for pandemic situations. Researchers have suggested various ways to stay informed about the situation (Cao et al, 2020). Natural Language Processing techniques are predominantly applied in the field of text analysis and text mining (Gohil et al, 2018). We are trying to incorporate similar techniques as those being used for healthcare and social media data, and to understand how efficient They are in analyzing the pandemic data. Data Mining is probably the best approach for wrangling data and generating some insights from it. Data mining can be further sub-divide into data visualization and data analysis. Visualization includes technique such as Exploratory Data Analysis (EDA) and text pre-processing. Analysis primarily implements techniques such as sentiment analysis, text analysis, polarity measurement of sentiments and topic mining/modeling. A study conducted by (Pokharel, 2020) has considered analysis of societal sentiments regarding Covid-19 thereby limiting the scalability and extraction of appropriate knowledge regarding the opinion and perspective of people towards the various covid-19 related aspects such as government policies, vaccination availability and dispersion of Covid-19 anxiety related to coronavirus across the world. Application of text analytic tools for information extraction has been a very active area of research. Researchers have done extensive analysis over text analytic tools (Iqbal et al, 2015), (Gu and Leroy, 2020) for information extraction. Building knowledge from information using healthcare data along with NLP has been conducted in the past (Savova et al, 2010). However, its application using textual data for pandemics is still an open research area. The objective of our research is to provide a text analytics based automated decision support tool which is capable of extracting the high-level informational features from raw data and categorizing them into knowledge dimensions. Furthermore, our approach can be treated as unified to deal with any traumatic and pandemic event related discussion on social media platform to understand the impact and effects. The remainder of the paper is organized as follows. In section 2 we will review the literature pertaining to Twitter and social medial data using the NLP tools for managing knowledge from Covid-19 pandemic Tweets. In the next section, we will conceptualize a knowledge extraction framework. Next, we will discuss our results and their implications. Finally, we will provide concluding remarks.

LITERATURE REVIEW

Widespread presence and enduring demand for the Internet has connected people more than ever, over time people are much more dependent on the Internet and social media to gather information and insight (Yang et al, 2020). Social media being such a powerful source sometime might be misleading and the chances of finding correct answers can be tricky. The amount of data available over social media platforms is huge and DISPERSED all over the internet. Generating meaningful insights or collecting accurate information is very difficult, especially when the topic is trending due to healthcare issues or a pandemic. Research conducted by (Cao et al, 2020) suggests that not all of the information available can be trusted and the quality of the information might be subject to revision, therefor generating insights from the vast amount of scattered data is highly complex.

Social media data especially Tweets from Twitter have generated some significant insights in finding answers related to healthcare and diseases. Research by (Gohil et al, 2018) suggests that analysis of social media data can provide important insight to health risks. However, application of these techniques for getting insights from pandemic is an active area of research. Research by (Carvalho and Harris, 2020) utilized pre-trained models to perform sentiment analysis and bag of words to study off-the-shelf technologies.

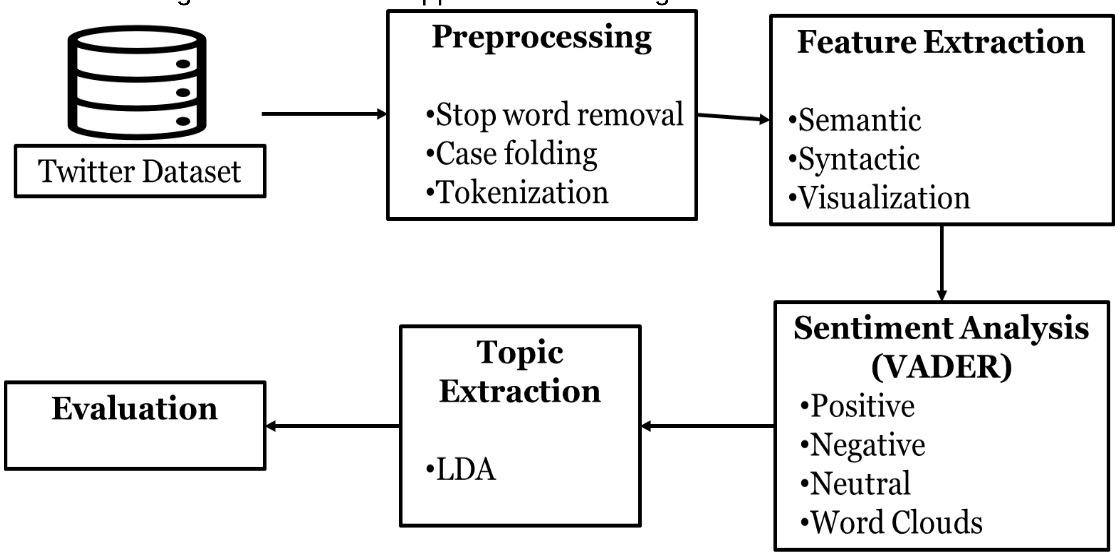
Text mining techniques using natural language processing has been extensively applied in healthcare domain (Basyal et al, 2020), (Chen et al, 2018). These techniques have been proven very effective in terms of generating insight either from free medical text, electronic health records or social media data. Text mining techniques such as sentiment analysis, and topic modeling can

provide valuable information. Topic modelling another popular technique provided promising results (Gu and Leroy, 2020).
Based on our review we can infer that a text mining approach using NLP is very beneficial for analyzing textual data. Twitter data on the other hand is a direct source of people's opinion available in textual form. Performing proper exploratory data analysis can reveal the type of techniques which could be applied for extracting knowledge. However, studies related to text mining have been largely confined to either text analyzation, exploratory study (Shahi et al, 2021) or sentiment analysis(Pokharel, 2020), to overcome this research gap we will apply major text pre-processing techniques such as stemming, stop word removal and tokenizing the text in order to prepare a large and quality corpus. In addition, we will perform text mining techniques (sentiment analysis and topic mining) capable of providing valuable insights. In this paper we propose a novel framework capable of processing raw data into information. To accomplish the objective, we will implement EDA to better understand the data. Further, we will conduct sentiment analysis to understand the common perception of users. Nest, we perform Latent Dirichlet Analysis (LDA) to extract the meaning in the form of topics. Finally, we consolidate the topics and formulate them in dimensions which will contribute towards better understanding of Covid-19 pandemic.

METHOD

We introduce a novel decision support framework for knowledge extraction using textual data (in our case Tweets). Our framework consists of six stages, capable of converting raw Tweet data into meaningful information if executed step by step. After importing the data, we will perform data preprocessing techniques such stop word removal and tokenization, this stage will provide us quality data for data analysis purposes. In the third stage we implement EDA using tools such as pycountry to extract the country name and use tools to extract the prominent features in the form of syntactic and semantic categories followed by visualizing them using bar graphs by location and trending topics. Sentiment analysis is performed in the fourth stage in order to better understand opinion and perception. The fifth stage consists of LDA where high level information is extracted in the form of topics which are then evaluated in the sixth stage to formulate knowledge dimensions.

Figure 1: Decision support and knowledge extraction framework.



Dataset

We collected publicly available Covid-19 Tweet data from Kaggle.com that comprises more than hundred and seventy thousand (179,108) tweets associated with #Covid-19. The dataset contains various user oriented (name, description, followers, friends), Tweet oriented (text, hashtags, sources) and location oriented (user location) information.

Data Preprocessing
Since a Tweet has a restriction of 280 characters (Vyas and El-Gayar, 2020), a sensible text data preprocessing is required over Tweets to eliminate unwanted stop words (i.e. , over, under, again, further, then, once, here, there) and redundant Tweets. Case folding such as lowering of Tweet content is important for further sentiment analysis. Moreover, tokenization is essential to create word tokens for feature extraction, vocabulary building and sentiment categorization. Basically, tokenization splits the Tweet sentences into words known as tokens for example, a "praying for good health" Tweet will be tokenized as praying, for, good and health tokens.

Feature Extraction

Semantic Features
Sentiment words are most frequently used features belongings to semantic categories that further indicate positive, negative, and neutral sentiments. (Giachanou and Crestani, 2016).

Syntactic Features
Unigrams, bigrams, and n-grams belong to syntactic feature categories and most often combine with semantic features (Giachanou and Crestani, 2016).

Sentiment Analysis

Polarity based sentiment analysis aim to analyze the text for detection of positive, negative and neutral sentiments. (Hutto & Gilbert, 2014.) developed valance aware dictionary for sentiment reasoning (VADER), that is a rule based model used for sentiment analysis on social media text that has the potential to perform better because of its sensitiveness to the sentiments in a social media context (Elbagir and Yang, 2019). VADER outperformed other state of the practice benchmark techniques such as Affective Norms for English Words (ANEW), Linguistic Inquiry and Word Count (LIWC) and the General Inquirer(Hutto & Gilbert, 2014.). Therefore, we utilized VADER to extract user sentiments from Covid-19 Tweets. An open-source Natural Language Toolkit (NLTK) library of python provides the VADER implementation to use as a tool for sentiment analysis.
Furthermore, the usage of word clouds to visualize word vocabularies spread among different sentimental categories makes the analysis more appealing and straightforward. Also, word clouds generated from a body of text can serve for a deeper analysis (Heimerl et al, 2014). Therefore, we utilized word cloud visualization for identified positive, negative, and neutral sentimental words categorized by VADER.

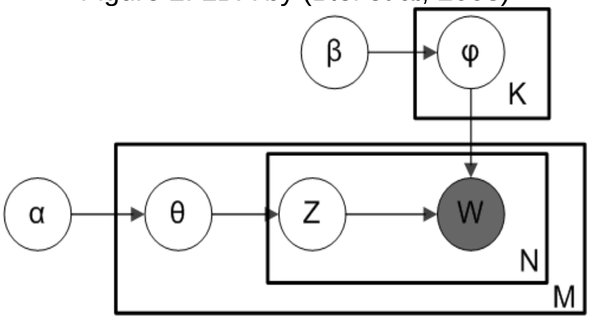
Topic Extraction

An unsupervised way of discovering and classifying a contextual information without having a prior understanding is known as topic modeling that is similar to clustering that extracts some group of items (topics). Creation of various themes and classification of text among those themes are two major aims of topic modeling. For example, a group of Tweet statements about corona, vaccine, vaccination, Covid-19 and medication can be grouped into a "corona vaccine" theme.

For the topic extraction from the Tweet, we utilized LDA technique. Where Latent represents the words and Dirichlet refers to probability distribution.

Latent Dirichlet Allocation (LDA)

Figure 2: LDA by (Blei et al, 2003)



The purpose of LDA is to explore the topics based on the words in the text document (Blei et al, 2003). LDA is an unsupervised learning technique that treat each document as an unordered bag of words. Primarily, LDA works on assumed topics based on a way the text document was created, then it associates the set of words related to assumed topics. As shown in Figure 2, K represents the assumed topics for the text document M. α represents the distribution of topics per document, β represents the distribution of words per topic, θ represents the distribution of topic for a document, φ represents the distribution of word for a topic. z represents the topic for the Nth word M and W represents the word. As a way to create a topic, LDA performs iterative steps. 1) Selection of number topics to be extracted from the documents. 2) traverse each document to assign the selected topic to each word of that document. 3) For each topic of each word in each document calculate the probability p(topic | document) and p(word | topic). 4) assign the new topic to a word by calculating probability p (topic | document) x p(word | topic). 5) iterate steps 3 and 4 until LDA determines similar topics for words in a document (Blei et al, 2003).

Evaluation

To evaluate the LDA topic model we utilized the perplexity and coherence as a measure. Perplexity measures how well LDA predicts the topics for unknown dataset i.e., held out test data. A lower perplexity score shows the better performance of LDA. Whereas coherence is similarity measure between high scoring words in a topic. A higher coherence score shows better model performance.

DISCUSSION AND RESULTS

Exploratory Data Analysis

An extensive EDA was performed to gain insights from the dataset to understand the behavior of Covid-19 related Tweet dispersion across the globe. Table 1 indicates the processing results performed on the Tweet location variable to find out exact total Tweet counts among top five locations. Before processing, the data possessed long named locations such as New York, USA we converted that into exact user location as a country name USA by utilizing pycountry python library. After analyzing Table 1 and figure 3 we observed that India and United States are the two major counties posting Covid-19 related Tweets. Figure 4 shows that approximately 32% of people utilized Twitter web and 22% of people used mobile applications to post their Tweet.

Table 1: Top Five Location by Tweet Count			
BEFORE PROCESSING		AFTER PROCESSING	
Location Name	Tweet count	Location Name	Tweet count
Unknown	36771	Unknown	36776
India	3741	India	15769
United States	2455	United States	9608
New Delhi, India	1721	Canada	6349
Mumbai, India	1401	Australia	2957

Figure 3: Top five Locations by Tweet count

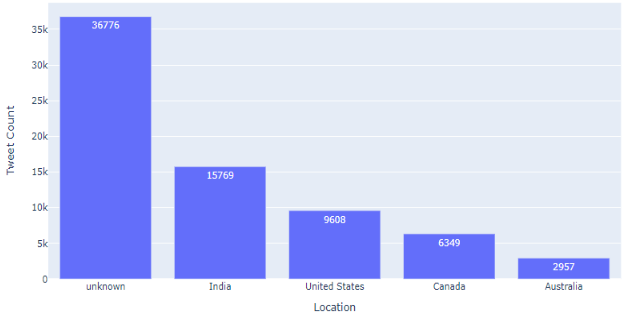
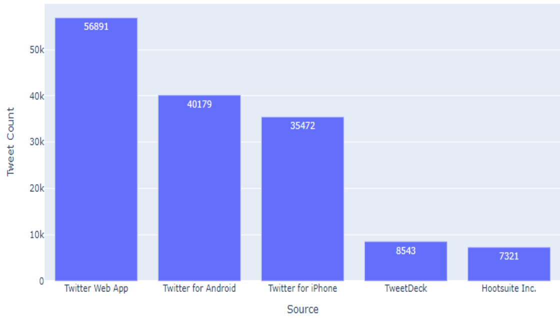


Figure 4: Top five Sources by Tweet count



Further as shown in figure 5 and 6 we analyzed the top trending hashtags (#) and mentions (@) utilized by people, respectively. #Covid-19 is the top trending hash tag used by the 26% of people and @realdonaltrum was the most mentioned Twitter profile followed by @WHO and @joebiden.

Figure 5: Top Trending #tags by Tweet count

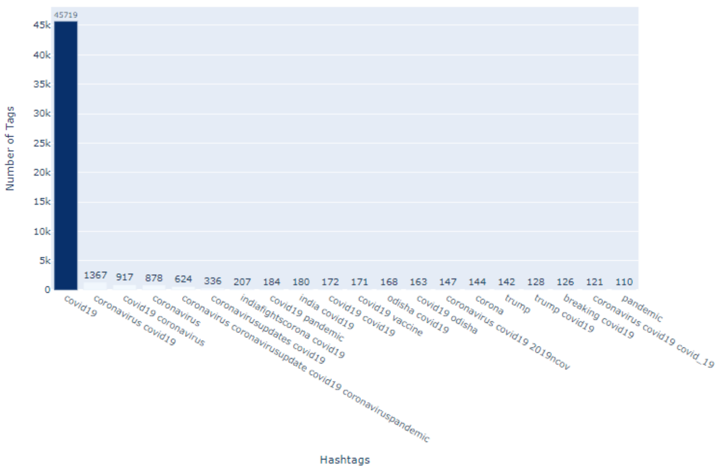
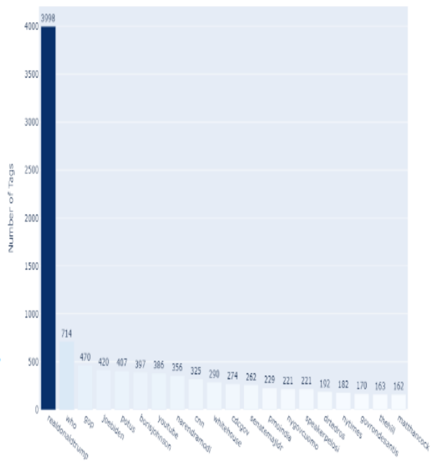


Figure 6: Top trending user mentions by Tweet count



We created a word cloud to analyze the most popular words in Tweet and hashtags used by users. As shown in figure 7, we observed that Covid-19, sanitizers, hand, stop, realDonaldTrump are the most used words in the Tweets. Whereas, coronavirus and Coiv1-9 were used as the majority terms of hashtags. Furthermore, after performing standard NLP techniques of text data preprocessing such as stop word removal, tokenization, and bag of words creation we generated the N-grams (sequence of words) of the Tweets to analyze the co-occurrence of the words that provides a tentative guess related to what users are trying to express. Since text in Tweet data is limited, we generated unigram, bigrams, and trigrams. As shown in Figure 8, we observed that generated N-grams disclosed important information related to user's feeling such as they mentioned about Covid-19 vaccines, tests, cases, and prevention measures.

Figure 7: Word cloud for popular buzzing words in tweet and hashtags

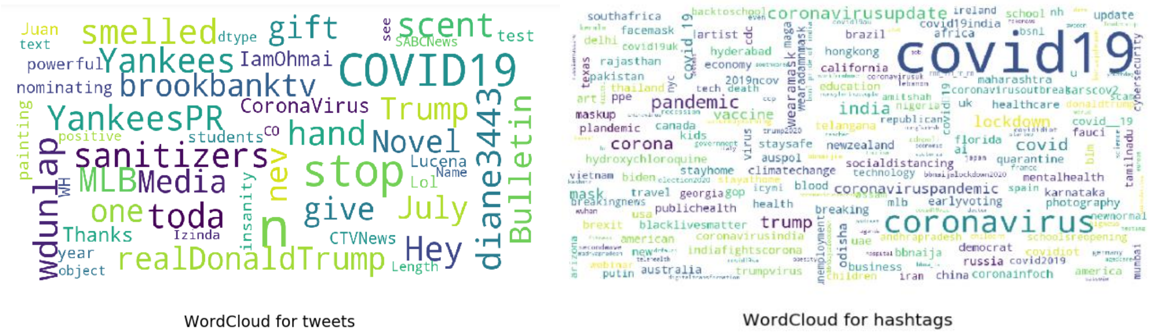
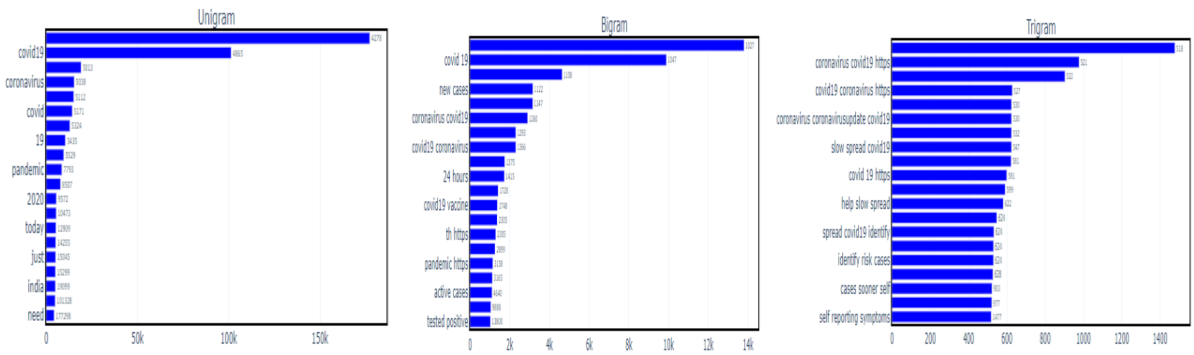


Figure 8: N-grams of tweets



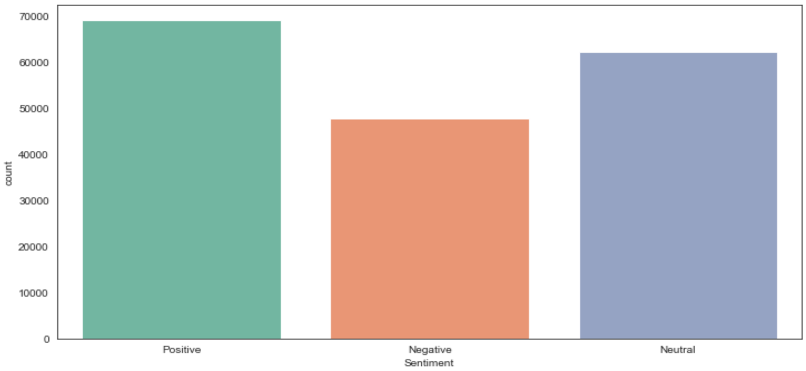
Sentiment Analysis

As shown in Table 2 and Figure 9, VADER generated the positive, negative, and neutral sentiments based on the text polarity. We used compound score as a threshold to categorize the Tweets, if the compound score >= 0.05 then Tweets are categorized as positive sentiment, if compound score <= -0.05 then Tweets are categorized as negative sentiment and if compound score is in between -0.05 and 0.05 then Tweets are categorized as neutral sentiment. We did not remove the emojis and emoticons because VADER also consider such indicators to generate compound+d score of polarity. As observed by the results 38.5% Tweets showed spread of positivity on Twitter followed by 34.7% neutral sentimental Tweets that further represents the optimistic perception of people towards the pandemic.

Table 2: Sentiment Distribution among Tweets

SENTIMENT	TWEET COUNT
Positive	69073
Neutral	62199
Negative	47836

Figure 9: Frequency of sentiments among Tweets.



Furthermore, we created word clouds for all three categories to examine most word occurrences in positive, neutral, and negative Tweets, respectively. As shown in Figure 10, Covid-19, people and case are among commonly dispersed high intensity words among all three categories. Positive sentiment word cloud showing some positive indicator words regarding Covid-19 awareness such as hope, safe, mask and support. Negative sentiment word cloud showing some negative indicators such as death, died, risk, infected and threat whereas neutral word cloud showing normal sentimental indicators such as want, need, help, support and safe. Although all categories have some similar sentimental indicators because of the Tweets' context.

Figure 10: Word cloud of Positive (green box), Negative (red box) and Neutral (yellow box) sentiment Tweets

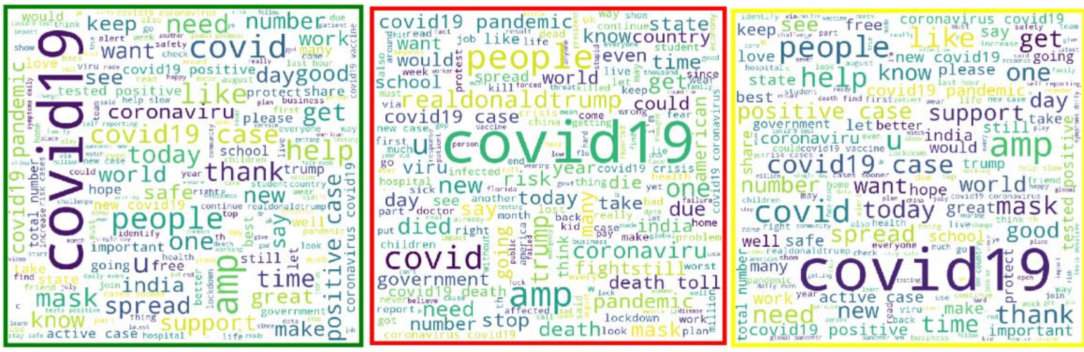


Table 3 shows the example of all three positive, negative and, neutral sentiment Tweets along with their polarity scores.

Table 3: Tweets Example Per Sentiment Category with their Polarity Scores.

SENTIMENT	EXAMPLE TWEET	SCORE
Positive	"GVMC sanitation staff carrying out the regular sanitation activities to keep the city clean and prevent the spread."	{'neg': 0.0, 'neu': 0.817, 'pos': 0.183, 'compound': 0.4215}
Negative	"By the time you see a spike in #Covid-19 cases, it is very hard to reverse it without drastic lockdown."	{'neg': 0.057, 'neu': 0.943, 'pos': 0.0, 'compound': -0.1027}

Neutral	“You now have to wear face coverings when out shopping - this includes a visit to your local Community Pharmacy also.”	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

DIMENSIONS OF COVID-19 TOPICS

We utilized the open source Gensim library to implement the unsupervised LDA topic modeling algorithm. Further we applied lemmatization over preprocessed Tweets with respect to similar nouns and adjectives because this provides more insights related to topics. Lemmatization is a process to eliminate inflectional ends from the word and return the dictionary form of words, such as vaccines converts into vaccine. We utilized pre-trained Spacy model named en_core_web_md that is an English multi-task CNN model trained on OntoNotes (annotated corpus of sources such as news, broadcast, talk shows, weblogs), with GloVe vectors for to assigns word vectors. We created the dictionary of 337,922 vocabulary size from the Tweet data to further create a document term matrix that explains the frequency of word among all Tweets for which we used doc2bow() function. Since Tweets have limited words, by utilizing high coherence scores we tuned LDA algorithm with number of topics, chunk size (number of Tweets in a batch/iteration), number of passes and number of iterations, alpha (Document-Topic Density) and eta (Word-Topic Density~ beta). Final optimal parameter values are, 15 number of topics, 500 chunk size, 50 passes, 50 iterations, 0.01 alpha and 0.9 eta. LDA model’s perplexity and coherence scores on Tweet data is -13.01 and 0.37 respectively. Table 4. shows the extracted topics along with their weightage scores and associated theme.

Table 4: Summary of Dimensions Formulated from the Extraction of Topics and Theme.

DIMENSIONS	THEME	CODE	TOPICS
Public leader (T2, T5, T9) School & college (T7, T3) Safety & Precaution measures (T2, T8, T10) Vaccine (T1) Economy Impact (T4)	Vaccine & time	T1	'0.104*"day" + 0.051*"year" + 0.035*"vaccine" + 0.031*"month" + ' '0.031*"first" + 0.029*"last" + 0.023*"covid" + 0.022*"bring" + ' '0.021*"spike" + 0.017*"old"
	Safety appeal	T2	'0.065*"covid" + 0.027*"thank" + 0.025*"give" + 0.025*"safe" + 0.022*"keep" ' '+ 0.021*"realdonaldtrump" + 0.019*"community" + 0.018*"life" + ' '0.018*"support" + 0.018*"fight"'
	Kids schooling	T3	'0.054*"covid" + 0.039*"school" + 0.036*"go" + 0.027*"come" + 0.026*"back" + ' '0.024*"know" + 0.022*"child" + 0.020*"want" + 0.019*"re" + 0.018*"amp"'
	Business Impact	T4	'0.086*"covid" + 0.051*"pandemic" + 0.038*"how" + 0.023*"business" + ' '0.019*"impact" + 0.016*"work" + 0.016*"crisis" + 0.015*"amp" + ' '0.012*"economy" + 0.011*"change"'
	Trump	T5	'0.054*"covid" + 0.038*"read" + 0.026*"trump" + 0.020*"check" + 0.017*"win" ' '+ 0.016*"more" + 0.015*"important" + 0.014*"article" + 0.014*"great" + ' '0.013*"story"'

Death (T6)	Death Report	T6	'0.140*"case" + 0.120*"covid" + 0.090*"new" + 0.075*"death" + 0.040*"report" ' '+ 0.039*"coronavirus" + 0.033*"total" + 0.031*"update" + 0.026*"number" + ' '0.019*"confirm"'
	Students	T7	'0.067*"covid" + 0.034*"virus" + 0.033*"student" + 0.032*"state" + ' '0.022*"country" + 0.017*"real" + 0.017*"travel" + 0.017*"medium" + ' '0.016*"outbreak" + 0.015*"coronavirus"'
	Lockdown & precaution	T8	'0.079*"covid" + 0.031*"lockdown" + 0.028*"care" + 0.028*"patient" + ' '0.023*"take" + 0.021*"continue" + 0.019*"hospital" + 0.013*"worker" + ' '0.013*"measure" + 0.012*"government"'
	COVID-19 testing & election	T9	'0.061*"covid" + 0.027*"state" + 0.019*"testing" + 0.013*"coronavirus" + 0.011*"top" + 0.010*"_" + 0.009*"contact" + 0.009*"early" + 0.008*"joebiden" + 0.008*"election"'
	Face mask	T10	'0.047*"covid" + 0.034*"mask" + 0.018*"wear" + 0.014*"amp" + 0.013*"how" + 0.012*"face" + 0.011*"give" + 0.010*"safe" + 0.010*"stay" + 0.010*"keep"'

CONCLUSION

In this research we have extensively applied state of the art text mining techniques. We proposed a novel framework for a decision support tool and knowledge extraction. Our research framework is capable of an end-to-end data analysis process, from data retrieval to knowledge representation. Our research converted the tacit knowledge available in the form or raw Tweets having no informational value into meaningful high-level information. The top-10 themes extracted in form of information have been categorized to create knowledge dimensions. Our research has at least two implications. First, the knowledge dimensions extracted from it will serve as a decision support tool for making informed decisions for pandemic situations especially during an era of misinformation. Our topic can serve as topic for further research, such as topic T10 (Face mask) can be researched for "mask mandate pre and post covid era". Secondly, our research will contribute to the knowledge base for researchers, those seeking to work in similar domain, such as generating the insights from COVID-19 pandemic using text mining. Our research demonstrated how we can extract knowledge from patterns of information found in raw data and represent them as topics grouped into dimensions.

REFERENCES

Basyal, G. P., Rimal, B. P., and Zeng, D. 2020. "A Systematic Review of Natural Language Processing for Knowledge Management in Healthcare," in Computer Science & Information Technology, AIRCC Publishing Corporation, July 11, pp. 275–285. (https://doi.org/10.5121/csit.2020.100921).

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," Journal of Machine Learning Research (3:Jan), pp. 993–1022. (https://www.jmlr.org/papers/v3/blei03a).

Cao, Q., Hu, W., Zhou, S., Luo, B., Yu, Y., Liao, S. S., Zhao, J. L., and Lau, R. Y.-K. 2020. "The Role of Media Coverage on Pandemic Containment: Empirical Analysis of the COVID-19 Case," PACIS 2020 Proceedings, p. 10.

Carvalho, A., and Harris, L. 2020. "Off-the-Shelf Technologies for Sentiment Analysis of Social Media Data: Two Empirical Studies," AMCIS 2020 Proceedings, p. 11.

Chen, X., Xie, H., Wang, F. L., Liu, Z., Xu, J., and Hao, T. 2018. "A Bibliometric Analysis of Natural Language Processing in Medical Research," BMC Medical Informatics and Decision Making (18), p. 14.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., and Scala, A. 2020. "The COVID-19 Social Media Infodemic," Scientific Reports (10:1), p. 16598. (https://doi.org/10.1038/s41598-020-73510-5).

Elbagir, S., and Yang, J. 2019. "Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment," Proceedings of the International MultiConference of Engineers and Computer Scientists, p. 5.

Ge, H., Wang, X., Yuan, X., Wang, C., Tianci, D., Yuan, Q., and Xiao, X. 2020. "The Epidemiology and Clinical Information about COVID-19," Eur J Clin Microbiol Infect Dis, p. 9.

Giachanou, A., and Crestani, F. 2016a. "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," ACM Computing Surveys (49:2), pp. 1–41. (https://doi.org/10.1145/2938640).

Gohil, S., Vuik, S., and Darzi, A. 2018. "Sentiment Analysis of Health Care Tweets: Review of the Methods Used," JMIR Public Health and Surveillance (4:2), p. e43. (https://doi.org/10.2196/publichealth.5789).

Gu, Y., and Leroy, G. 2020. "Large-Scale Analysis of Free-Text Data for Mental Health Surveillance with Topic Modelling," in Association for Information Systems, , August, p. 6.

Guo, Y.-R., Cao, Q.-D., Hong, Z.-S., Tan, Y.-Y., Chen, S.-D., Jin, H.-J., Tan, K.-S., Wang, D.-Y., and Yan, Y. 2020. "The Origin, Transmission and Clinical Therapies on Coronavirus Disease 2019 (COVID-19) Outbreak – an Update on the Status," Military Medical Research (7:1), p. 11. (https://doi.org/10.1186/s40779-020-00240-0).

Heimerl, F., Lohmann, S., Lange, S., and Ertl, T. 2014. "Word Cloud Explorer: Text Analytics Based on Word Clouds," in 2014 47th Hawaii International Conference on System Sciences, , January, pp. 1833–1842. (https://doi.org/10.1109/HICSS.2014.231).

Hutto, C. J., and Gilbert, E. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Association for the Advancement of Artificial Intelligence., p. 10.

Iqbal, E., Mallah, R., Jackson, R. G., Ball, M., Ibrahim, Z. M., Broadbent, M., Dzahini, O., Stewart, R., Johnston, C., and Dobson, R. J. B. 2015. "Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register," PLOS ONE (10:8), (C. A. Ouzounis, ed.), p. e0134208. (https://doi.org/10.1371/journal.pone.0134208).

Kumar, A., and Sebastian, T. M. 2012. "Sentiment Analysis on Twitter," International Journal of Computer Science IJCSI (9:4), p. 7.

Mehta, P., McAuley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., and Manson, J. J. 2020. "COVID-19: Consider Cytokine Storm Syndromes and Immunosuppression," The Lancet (395:10229), pp. 1033–1034. (https://doi.org/10.1016/S0140-6736(20)30628-0).

Pokharel, B. P. 2020. "Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal," SSRN Electronic Journal. (https://doi.org/10.2139/ssrn.3624719).

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., and Chute, C. G. 2010. "Mayo Clinical Text Analysis and Knowledge Extraction System (CTAKES): Architecture, Component Evaluation and Applications," Journal of the American Medical Informatics Association (17:5), pp. 507–513. (https://doi.org/10.1136/jamia.2009.001560).

Shahi, G. K., Dirkson, A., and Majchrzak, T. A. 2021. "An Exploratory Study of COVID-19 Misinformation on Twitter," Online Social Networks and Media, p. 100104. (https://doi.org/10.1016/j.osnem.2020.100104).

Vyas, P., and El-Gayar, O. 2020. "Credibility Analysis of News on Twitter Using LSTM: An Exploratory Study," AMCIS 2020 Proceedings. (https://aisel.aisnet.org/amcis2020/social_computing/social_computing/17).

Yang, Z., Ram, S., and Currim, F. 2020. "What Makes an Online Suggestion A Good One for Online Health Communities," AMCIS 2020 Proceedings, p. 11.

DECISION SCIENCES INSTITUTE

Utilize Digital Platforms for Business Partners Selection to Achieve Competitive Advantage and Platform-based Satisfaction

My-Trinh Bui
VNU International School - Vietnam National University, Hanoi
trinhbm@isvnu.vn

Huong-Linh Le
VNU International School - Vietnam National University, Hanoi
linhlh@isvnu.vn

ABSTRACT

Digital platforms have substantially reinvented enterprises to improve their performance, exploit existing markets more effectively, and enter new markets. This study aims to analyze the selection of business partners to achieve competitive advantage by using digital platforms in the context of Covid-19 pandemic. We collected and analyze the data of 355 enterprises using the theories of business partner selection, service delivery co-innovation, and perspectives of competitive advantage. Digital platforms applications are proved to be critical in the global partner selection process, especially in the present context, when all countries are affected by the Covid-19 epidemic.

KEYWORDS: Business Partners, Digital Platforms, Co-innovation, Competitive Advantage, Pandemic

INTRODUCTION

In a traditional way, enterprises typically initiate a marketing campaign or have sales representatives' approach potential business partners such as customers or suppliers. However, selecting a partner in this manner presents a number of challenges such as geographic distance and complex legal procedures. With today's rapid technological advancements, digital technology platforms application is becoming increasingly important and urgent for businesses (Costa et al., 2020; Jin & Hurd, 2018; Ruggieri et al., 2018). A digital platform enables a focal firm to generate, deliver, and acquire value, while business partners selection is a critical co-innovation process which is related with radical business transformation in the structure of an existing relationship pattern, aimed at gaining win-win competitive advantages.

The new market landscape heavily affected by Covid-19 since late 2019 with travel constraints between nations, supply chain disruptions, and quickly shifting consumer demand forces firms to change their attitude towards digital transformation in a faster manner. First, firms have to rely mostly on the digital platforms to find their potential and strategic partners in the market shaken by the pandemic. Second, they need to find how to co-innovate new service distribution processes with their business partners. Third, firms have to determine how to evaluate the feasibility of co-innovation in service delivery during the Covid-19 pandemic. Finally, the judgment of firms towards the chosen digital platforms is still unclear.

This study aims to gain four main objectives. First, the paper explores how using digital platforms helps businesses find their partners in the Covid-19 period. Second, it investigates how companies collaborate with their business partners to develop alternative service delivery processes in the event of a pandemic. Third, the study examines the effectiveness of service delivery co-innovation during the Covid-19 pandemic in terms of competitive advantage. Finally, it clarifies the firms' satisfaction level with the application of digital platforms in choosing their business partners during the Covid-19 pandemic.

LITERATURE REVIEW

There are numerous studies conducted to find out about enterprises' difficulties when entering new markets. In a microeconomic view, the obstacles for firms to enter a new market are liability of foreignness, psychic distance (Morrison, 2011), financial limitations, and lack of experience (Scott-Kennel, 2013).

Finding Business Partners in New Markets

Organizational characteristics such as ownership form, scale, and age, which vary between domestic and international companies, can grant differentiated local advantages (Bellak, 2004). Digital platforms might help enterprises to get over resource constraints in financial capitals (Jin & Hurd, 2018) and support firms to understand more about the market. But what is known about popular platforms (both digital and physical) that the firms can meet and see each other's cut out for co-innovation during the Covid-19 pandemic is still limited.

Co-innovation by Cooperating with Local Partners

Traditional definitions of co-innovation place a heavy emphasis on third-party engagement, such as a consumer or a supplier with a business relationship (Dawson et al., 2014; Yeniyurt et al., 2014). Selecting suitable business partners is one of the obstacles facing companies that co-innovate with them (Yeniyurt et al., 2014). This question remains unanswered in studies on how to choose suitable business partners that co-innovate new products/services, especially in the context of the Covid-19 pandemic. Using the theory of partner selection (Emden et al., 2006), this study finds out: (i) how companies select business partner to have competitive advantage in market using digital platforms during the Covid-19 pandemic, and (ii) how they use digital platforms to collaborate with their partners on innovative service delivery processes in order to combat the pandemic's challenges.

Importance of Digital Platforms for Co-innovation during COVID-19 Pandemic

In several industries, companies can achieve core competencies, synergies, shared learning and information sharing, cost reductions by finding business partners (Kozarkiewicz & Kabalska, 2016). Platforms can be used in several ways, according to an increasing body of literature in innovation management (de Reuver et al., 2018; Gawer & Cusumano, 2014; Karhu & Ritala, 2020). The Covid-19 pandemic is forcing businesses to collaborate in new directions, resulting in ecosystem-wide innovation (Accenture, 2020; Deloitte, 2020; KPMG, 2020; McKinsey, 2020). The aim of this study is, therefore, to leverage the potential of innovative collaboration in a business sector via digital platforms in the Covid-19 pandemic.

THEORETICAL FRAMEWORK

Business Partners Selection

During this time, resource dependence theory has been widely used in studies to illustrate how organizations' minimize environmental interdependence and volatility (Hill & Hellriegel, 1994; Hillman et al., 2009). Emden et al. (2006) suggests three core factors to ensure an efficient partnership process: relational alignment, technical alignment, and strategic alignment (Tsou et al., 2015). First, experience or "partner expertise" is an important element in cooperation (Auh et al., 2007; Lusch et al., 2007; Subramani & Venkatraman, 2003). Second, resource complementarity is another part of technical alignment (Queiroz et al., 2020). Only by combining of expertise and resources or their "partner complementary" will firms be able to maximize or build opportunities (Emden et al., 2006; Hill & Hellriegel, 1994; Luo, 2002). Finally, according to Selnes & Gønhaug (2000), "partner reliability" is a crucial factor in establishing collaborative partnerships dependent on commitment. As a result, three requirements for choosing a business partner emerge: Partner expertise, Partner complementarity, and Partner reliability.

Service Delivery Co-innovation

Service delivery refers to the distribution of a service (Zeithaml et al., 2002) or a good (Lovelock & Wright, 2001; Moorman & Rust, 1999) to consumers. Service delivery must be planned, developed, optimized, and tested to satisfy the needs of customers (Scheuing & Johnson, 1989). Co-innovation can be aided by a sequence of interactions, such as *communication encounters, usage encounters, and service encounters,* among others (Payne et al., 2008).
New communication encounters (NCE) are events aimed at connecting with consumers and promoting and propelling dialogues. Besides country-level collaboration platforms, firms also can deliberately provide updated information to customers via firm-level platforms such as Internet portals, keyword ads, service launch ceremonies, and online social networking platforms such as Linkedin, WhatsApp, Zalo, Skype, Facebook, Line, WeChat, Viber, Yahoo Messenger, Google +, and so on.
New usage encounters (NUE) are encounters that expand on a process that a client uses for a product or service to provide services that facilitate use, such as self-service technologies. One example is mobile banking.
New service encounters (NSE) is defined as ways that customers engage with customer support staff or service applications. Service men and women, for example, offer services such as videoconferences or post-sales support such as 24-hour call centers.
Co-innovation is characterized as an innovation co-created by a collaboration between firms, suppliers, and customers for the purpose of sharing information, costs, and benefits in order to generate unique value for end users (Dawson et al., 2014). Thus, co-innovation entails cooperation with a variety of strategic partners which can be carried out in a variety of ways, based on how companies' technical skills are supplemented by those of their strategic partners (Baldwin & Von Hippel, 2011). In this vein, co-innovation in service distribution can be viewed as an innovative mode of delivery that provides customers with additional convenience. Furthermore, as previously stated, service delivery is a sequence of two-way encounters. As a result, *service delivery co-innovation* is described specifically as a form of encounters that involves New communication encounters, New usage encounters, and New service encounters.

Competitive Advantage

Competitive advantage is gained by completely deploying and using exclusive, useful, and unique tools and skills, which can be interpreted externally as result success and internally as operational capabilities (Bhatt & Grover, 2005; Tian et al., 2010). Thus, companies can gain

greater competitive advantage in a variety of ways, including creativity, cost-cutting, market-based, employee-based, or relationship-based competitive advantages (Barney, 1995; Dyer & Singh, 1998; H. C. Wang et al., 2009). We use two competitive advantages to characterize the efficacy of service delivery co-innovation: *market-based competitive advantage*, which is gained via the development of unique products and services in markets, and *employee-based competitive advantage*, which is gained via the development of workers' distinctive skills and capabilities.
Organizational inputs affect organizational efficiency completely by organizational systems, according to the *input–process–output* paradigm (Deeter-Schmelz & Ramsey, 2003). This research employs this model to connect three variables: business partner selection (input variable), co-innovation in service delivery (process variable), and competitive advantage (output variable) (see Figure 1). Business partner selection includes partner reliability, complementarity, and expertise. Competitive advantage encompasses market-based competitive advantage and employee-based competitive advantage. Service delivery co-innovation includes New communication encounters (NCE), New usage encounters (NUE), and New service encounters (NSE). In the next segment, we'll look at how business partner selection affects service delivery co-innovation, which leads to a competitive advantage for companies.

HYPOTHESIS

Partner Reliability

A critical element in developing reliability is the ability to build inter-organizational partnerships and allegiance, as well as the ability to sustain and retain commitment (Selnes & Gønhaug, 2000). The value of a partner's trustworthiness is apparent because an untrustworthy partner can supply false information, exposing a business to needless cost (McKnight et al., 2017). Partner reliability is described in this study as the ability to control working environments efficiently and consistently, as well as participate in mutual cooperative activity. Trust improves stability, which improves productivity and value development, especially in co-innovation (Jayashankar et al., 2018). Thus, businesses that want service delivery co-innovation as a tactic must be able to gain partners' confidence and commitment in order for a collaboration solution to succeed.

H1 a) Partner reliability has a positive effect on NUE
b) Partner reliability has a positive effect on NSE
c) Partner reliability has a positive effect on NCE

(NUE = new usage encounters, NSE = new service encounters, NCE = new communication encounters.)

Partner Complementarity

Partner complementarity, described as a combination of partners' distinct competencies (Büyüközkan & Görener, 2015), is the most important factor to remember when choosing a partner. Complementarity refers to how distinct, interdependent, and mutually supportive two firms' resources are (L. Wang & Zajac, 2007). Partner complementarity indicates that a firm's capital may be offset by the resources of a partner (Powell & Dent-Micallef, 1997; Tippins & Sohi, 2003; Wade & Hulland, 2004). As a result, collaborators have a certain area of control that can compensate for the shortcomings of their potential cooperative partners. Complementary collaborators have unique tools, expertise, knowledge, and other advantages that facilitate cooperation (Cummings & Holmberg, 2012; Furlotti & Soda, 2018). Spin (2012) backs up this point, claiming that collaborators with complementary routines have advanced stages of service delivery co-innovation.

H2 a) Partner complementarity has a positive effect on NUE
b) Partner complementarity has a positive effect on NSE
c) Partner complementarity has a positive effect on NCE

Partner Expertise

Firms also seek partners with expertise and strengths from which they can benefit in order to improve their competitive advantage (Hitt et al., 2000). Partner expertise shows the partners' abilities or competencies; for example, a skilled nursing partner will provide accurate and timely information to improve service transactions (Kelley et al., 1990; Perry et al., 2019). The degree to which either partner is capable of supplying a service or providing necessary data linked to a related underlying technology can be replicated by partner experience (Bendapudi & Berry, 1997; Parmigiani, 2007). Firms who cooperate could improve the effectiveness of their offerings by combining their partners' expertise and skills. Partner expertise is described in this analysis as partners' specialized knowledge and skills, as well as their ability to provide accurate data. Co-innovation is the process of generating new ideas, skills, resources, and prospects across organizational boundaries (Snow et al., 2005). Though a partner's high level of expertise can lead to inflexible, egotistic behavior that stifles creative efforts, it can be argued that this partner's knowledge, skills, and abilities can provide a source of useful inputs for the creation and growth of new ideas.

H3 a) Partner expertise has a positive effect on NUE
b) Partner expertise has a positive effect on NSE
c) Partner expertise has a positive effect on NCE

Co-innovation and Firms’ Competitive Advantage Section

According to previous research, service delivery creativity is a key determinant of a company's competitive advantage (Chen et al., 2009; Tsou et al., 2015). Following this logic, we agree that co-innovation in service delivery will help companies achieve a competitive advantage. Creative approaches that result in innovative service delivery systems are referred to as co-innovation in service delivery. The three main events are, as previously said, new communication encounters, new usage encounters, and new service encounters. As a result, new communication, usage, and service encounters are likely to have an effect on market-based competitive advantages and employee-based competitive advantage differently (Maryono et al., 2019). This study aims to develop a comprehensive integrated model which helps in explaining the impact of value chain integration’s (VCI)’s and relational capability impact on co-innovation in a network and competitive advantage.
Managers will use these new usage methods to select the right methods for fully modifying or enhancing their current service delivery structures. Furthermore, innovative service management mechanisms will help staff develop their expertise and strengths in communicating with clients, resulting in improved service offers. As a result, workers may have superior knowledge, expertise, and talents that are both exceptional and impossible to replicate (Lee et al., 2012).

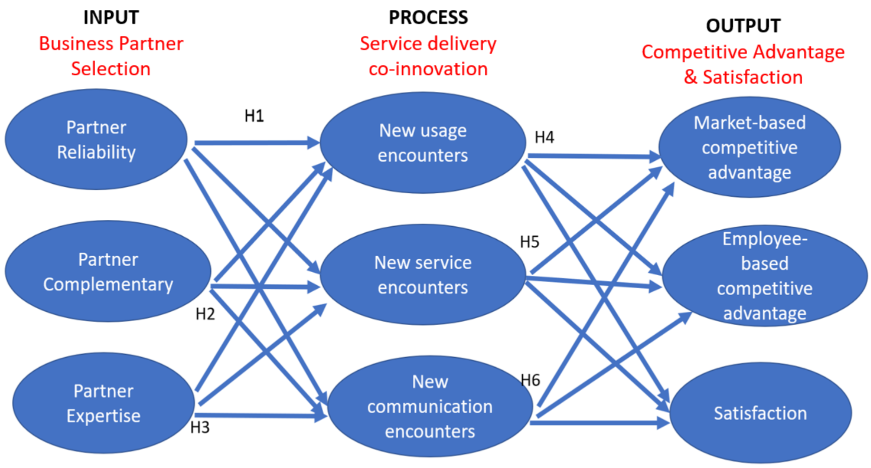
Satisfaction – It is self-developed construct designed by the research team. There were several research showing satisfactions with the digital platform development between partners can increase competitive advantage and sustainability. Public and private sector players frequently enter into relationships with ambiguous perceptions of the advantages to be achieved, but once active, they are generally happy with the outcomes (Hartwich et al., 2004). Given the high level of success that partners have with public-private collaborations, they are a promising new growth tool.

H4: NUE has positive effects on (a) market-based competitive advantage (b) employee-based competitive advantage and (c) satisfaction

H5: NSE has positive effects on (a) market-based competitive advantage (b) employee-based competitive advantage and (c) satisfaction

H6: NCE has positive effects on (a) market-based competitive advantage (b) employee-based competitive advantage and (c) satisfaction

Figure 1: The conceptual model



RESEARCH METHODOLOGY

Data Collection and Sample

We conducted a survey of global entrepreneurs from a wide range of sectors. Total of the 335 samples were selected primarily based on Linkedin professional business network of a global diversified business group located in Vietnam. They have established their own business networking platforms via Linkedin to find the partnerships for their state-of-the-art newly-constructed business.
The informants come from a variety of backgrounds, ranging from front-line employees, supervisors to general managers. A total of 335 questionnaires were completed and returned. An organization's data for independent and dependent variables was provided by a single key informant, which may be a source of widespread system bias. We have used the smart PLS version 3 approach to determine the strength of any possible traditional process bias.

Table 1 shows demographic information for individuals of each company.

Table 1: Demographics

Measure	Item	Frequency	Percent
Gender	Male	215	64.2
	Female	120	35.8
Marital Status	Do not disclose	1	0.3
	Married	249	74.3
	Divorced/Single	85	25.4
Age	Under 20	4	1.2
	21- 40	279	83.2
	41- 59	44	13.2
	>=60	8	2.4
Income (USD)	Under S20.000	52	15.5
	\$20.000 - \$44.999	48	14.3
	\$45.000 - \$139.000	13	3.9
	\$l40.000 - \$l 49.999	103	30.7
	\$l50.000 - \$l 39.999	119	35.5
Education	Collage & University	216	64.5
	Graduate School	114	34
	Jurior highschool & Vocation	5	1.5
Position	Front-line employee	28	8.36
	General manager	46	13.73
	Department manager	54	16.12
	Outlet manager	16	4.78
	Specialist/Expert	27	8.06
	Supervisor	67	20.00
	Senior workers	93	27.76
	Others	4	1.19
Professional Tenure	<6 years	172	51.34
	6-12 year	109	32.54
	>12 years	1	0.30
Firm size	<\$1 million	83	24.8
	\$1-10 millions	90	26.9
	\$10-20 millions	72	21.5
	\$20-50 millions	47	14.0
	\$50-100 millions	27	8.1
	> \$100 millions	16	4.8
Employee Size	< 50 people	136	40.6

	50-249	139	41.5
	> 249 people	60	17.9

Operational Definition of the observed variables representing the used terms in business. The requirements for selecting partners differ by organization, depending on the industry and the objectives that must be met when cooperating. The below are the three major criteria: partner reliability, partner complementarity, and partner expertise.
The scale of resource complementarity on the desire to accomplish the target and the enhanced benefit given by another firm (Chung et al., 2000; Tippins & Sohi, 2003; Wade & Hulland, 2004) was modified to assess *partner reliability* on seven items.
Partner complementarity refers to the fact that the two partner firms do not equally add capital to the partnership in both form and quantity as well as the potentials and skills of the two parties.
Partner expertise is measured by market analysis ability, ability to build strategy and credibility which allow a business to offer accurate, reliable and appropriate service and product details to their partner, improve each company's expertise.
Co-innovation in service delivery is based on two criteria: new usage encounters and new service encounters. New usage encounters or experiences are new approaches that customers use for a product or service, as well as the services that enable usage. New service encounters mean new ways that help customers interact with customer service staff or service applications most conveniently and simply.
The *New communication encounters* (NCE) factor also greatly reinforces the above factors. Companies can completely provide information through new digital platforms (Website, Facebook, Zalo, etc).
Finding a suitable partner for co-innovation is a springboard for businesses to gain competitive advantage. The principles for determining competitive advantage are market-based and employee-based competitive advantage. Market-based competitive advantage means in a new market, the firms satisfy customers' needs, providing better service quality than competitors. Employee-based competitive advantage means increasing staff job satisfaction, enhancing staff experience and domain knowledge, uplifting staff innovative capability.

Table 2: Descriptive Statistics and PLS-CFA Results

	SFL	**Cronbach's Alpha**	**rho_A**	**CR**	**AVE**	**VIF**	**Mean**	**SD**
ECA1	0.884	0.796	0.799	0.881	0.712	2.113	5.555	1.190
ECA2	0.788					1.440	5.549	1.196
ECA3	0.856					1.969	5.597	1.165
MCA1	0.856	0.752	0.756	0.859	0.670	1.729	5.460	1.206
MCA2	0.764					1.352	5.579	1.125
MCA3	0.833					1.630	5.600	1.152
NCE1	0.869	0.685	0.685	0.864	0.760	1.373	5.657	1.108
NCE3	0.876					1.373	5.666	1.057
NSE1	0.818	0.691	0.693	0.829	0.619	1.472	5.693	1.081
NSE2	0.733					1.218	5.699	1.107
NSE3	0.807					1.489	5.633	1.092
NUE1	0.811	0.790	0.793	0.864	0.615	1.782	5.621	1.186
NUE2	0.753					1.586	5.788	1.093
NUE3	0.746					1.521	5.624	1.099

NUE4	0.824					1.864	5.713	1.096
PC1	0.832	0.817	0.826	0.872	0.579	1.982	5.696	1.145
PC3	0.809					1.913	5.746	1.111
PC4	0.700					1.480	5.699	1.063
PC5	0.752					1.635	5.636	1.127
PC6	0.701					1.489	5.588	1.233
PE1	0.763	0.822	0.823	0.876	0.585	1.765	5.675	1.135
PE3	0.807					1.922	5.725	1.049
PE4	0.731					1.799	5.719	1.010
PE5	0.766					1.693	5.579	1.153
PE6	0.756					1.929	5.716	1.079
PR2	0.726	0.835	0.836	0.879	0.549	1.694	5.794	1.061
PR3	0.709					1.546	5.875	1.046
PR4	0.764					1.801	5.752	1.088
PR5	0.782					1.919	5.737	1.137
PR6	0.704					1.517	5.624	1.215
PR7	0.756					1.824	5.746	1.140
SAT1	0.760	0.805	0.808	0.865	0.562	1.629	5.672	1.199
SAT2	0.759					1.678	5.663	1.137
SAT3	0.758					1.663	5.627	1.154
SAT4	0.704					1.638	5.499	1.119
SAT5	0.765					1.626	5.537	1.191

SFL= Standardized Factor Loading, CR= Composite Reliability, Average Variance Extracted (AVE), SD= Standard Deviation, VIF= variance inflation factor, NUE = new usage encounters, NSE = new service encounters, NCE = new communication encounters, PR=Partner Reliability, PC=Partner Complementary, PE=Partner Expertise, MCA=Market-based competitive advantage, ECA=Employee-based competitive advantage, SAT=Satisfaction

Item PR1, PE2, PC2, NCE2 have been removed to enhance the validity and reliability of the model. In Table 3, all of our hypothesis were supported.

Table 3 : PLS Coefficient Path Analysis

	Path coefficient	**Standard Deviation (STDEV)**	**T Statistics**	**P Values**	**VIF**
PR -> NUE (H1.a)	0.298	0.085	3.509	0.000	3.547
PR -> NSE (H1.b)	0.214	0.079	2.710	0.007	3.547
PR -> NCE (H1.c)	0.191	0.099	1.918	0.056	3.764
PC -> NUE (H2.a)	0.222	0.078	2.834	0.005	4.143
PC -> NSE (H2.b)	0.177	0.069	2.556	0.011	4.143
PC -> NCE (H2c)	0.152	0.050	2.058	0.004	4.730
PE -> NUE (H3.a)	0.380	0.098	3.875	0.000	4.634
PE -> NSE (H3.b)	0.505	0.071	7.143	0.000	4.634

PE -> NCE (H3.c)	0.274	0.093	2.937	0.003	4.219
NUE -> MCA (H4a)	0.347	0.061	5.710	0.000	2.648
NUE -> ECA (H4b)	0.263	0.114	2.318	0.021	2.648
NUE -> SAT (H4c)	0.391	0.056	6.978	0.000	2.648
NSE -> MCA (H5a)	0.412	0.069	5.969	0.000	3.155
NSE -> ECA (H5b)	0.411	0.101	4.079	0.000	3.155
NSE -> SAT (H5c)	0.311	0.076	4.115	0.000	3.155
NCE -> MCA (H6a)	0.136	0.048	2.825	0.005	2.228
NCE -> ECA (H6b)	0.160	0.078	2.048	0.041	2.228
NCE -> SAT (H6c)	0.202	0.054	3.723	0.000	2.228

RESULTS AND DISCUSSION

Discussion

This research explore how companies choose business partners by using digital platforms to foster co-innovation, gain competitive advantage, and satisfaction during the Covid-19 pandemic. Selecting business partners has a beneficial impact on service delivery co-innovation, which is linked to market-based and employee-based competitive advantages. This supports the idea that businesses should use digital platforms to co-operate and grow in order to achieve business innovation.

Competitive advantage and satisfaction may become critical factors during the progress of service delivery co-innovation. Through the influence of service delivery co-innovation, partner reliability, partner complementarity, and partner expertise have a favorable impact on both market and employee-based competitive advantages. In specific, service delivery co-innovation is a strategic innovation mechanism that refers to the introduction of innovative products and services. Firms who can gain a competitive advantage by taking into account the views and knowledge of their business partners. This study introduces new communication encounters, new usage encounters, and new service encounters to existing or new customer groups to help firms to be empowered to achieve service delivery co-innovation, leading to market-based and employee-based competitive advantage.

Theoretical Implications

First, this research adds to the documentation of service creativity by integrating service delivery trends. This study indicates that the components of service delivery co-innovation including new communication experiences, new usage encounters, and new service encounters. As a result, this research represents a step toward a more realistic and evolved form of service innovation.

Second, based on the input–process–output model, our analysis emphasizes an interconnected perspective to link business partner selection, service delivery co-innovation, and competitive advantage (Deeter-Schmelz & Ramsey, 2003). In this study, service delivery co-innovation is characterized as a key mechanism that defines a company's competitive advantage.

Third, this research adds to partner selection literature by tailoring partner selection theory (Emden et al., 2006), which describes how such partner awards a sustainable sector service co-innovation, to balance the technical (partner competence and partner complementarity), and strategical (partner reliability) alignment theory.

Fourth, we emphasize two aspects of competitive advantage: business and employee-based. We suggest that competitive advantages on the markets and personnel are relevant as they demonstrate how businesses continually grow their partner ships to form their co-innovations in the distribution of services.

Managerial Implications

Companies should consider creating a membership model that allows members to establish reliable, complementary, and expertise partnership through sharing platforms. Platforms should conduct external trend research so that it can provide related, revenue-generating services. Companies should commit more resources to similar service delivery systems.
Second, the results of this study show that a firm's initial investment, especially in selecting partners with complementary expertise (partner complementarity and expertise), determines how much it contributes to service delivery co-innovation. Since customization markets are closely related to co-innovation, the company and the client can choose each other based on their mutual preferences (partner complementarity) and collaborate (partner reliability and expertise) to achieve a creative end result.
Third, diversity is an essential engine of co- innovation and a crucial part of global growth. Senior executives are discovering that creativity and the advancement of innovative innovations involve a complex range of viewpoints, insights, and backgrounds. The findings have important consequences for businesses in terms of the skills needed to leverage external expertise for innovation (Goyal et al., 2020).
Fourth, co-innovation in service management has a different impact on a firm's market-based competitive advantage and employee-based competitive advantage (competitive advantage). A company's competitive advantage needs differentiation based on co-innovation in service delivery. Businesses should invest in improving their employees' skills and capabilities.
Fifth, the SARS-CoV-2 coronavirus has sparked a burst of pace to respond to shifting conditions and demands. The pandemic has shown the importance of agility and accelerated creativity. Companies who want to improve their ability to innovate easily keep investing in digital transformation.

Limitations and Future Research

There are some deficiencies in this study that may lead to new research directions. First, we did not explore the different ways of co-innovating service providers. Second, our sample should generate the cultural influence and industry diversification on international business partner. In addition, it also depends heavily on the technological background as well as the culture of each company and each industry.

REFERENCES

Accenture. (2020). *COVID-19: Post-Coronavirus technology trends.* Retrieved from https://www.accenture.com/cz-en/insights/technology/tech-vision-coronavirus-trends, May 14, 2021.
Auh, S., Bell, S. J., McLeod, C. S., & Shih, E. (2007). Co-production and customer loyalty in financial services. *Journal of retailing, 83*(3), 359-370.
Baldwin, C., & Von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization science, 22*(6), 1399-1417.
Barney, J. B. (1995). Looking inside for competitive advantage. *Academy of Management Perspectives, 9*(4), 49-61.

Bellak, C. (2004). How domestic and foreign firms differ and why does it matter? *Journal of economic surveys, 18*(4), 483-514.
Bendapudi, N., & Berry, L. L. (1997). Customers' motivations for maintaining relationships with service providers. *Journal of retailing, 73*(1), 15-37.
Bhatt, G. D., & Grover, V. (2005). Types of information technology capabilities and their role in competitive advantage: An empirical study. *Journal of management information systems, 22*(2), 253-277.
Buchanan, L. (1992). Vertical trade relationships: the role of dependence and symmetry in attaining organizational goals. *Journal of Marketing Research, 29*(1), 65-75.
Büyüközkan, G., & Görener, A. (2015). Evaluation of product development partners using an integrated AHP-VIKOR model. *Kybernetes*.
Cenamor, J., Parida, V., Oghazi, P., Pesämaa, O., & Wincent, J. (2019). Addressing dual embeddedness: The roles of absorptive capacity and appropriability mechanisms in subsidiary performance. *Industrial Marketing Management, 78*, 239-249.
Chen, Y.-S., Lin, M.-J. J., & Chang, C.-H. (2009). The positive effects of relationship learning and absorptive capacity on innovation performance and competitive advantage in industrial markets. *Industrial Marketing Management, 38*(2), 152-158.
Chung, S., Singh, H., & Lee, K. (2000). Complementarity, status similarity and social capital as drivers of alliance formation. *Strategic Management Journal, 21*(1), 1-22.
Costa, E., Soares, A. L., & de Sousa, J. P. (2020). Industrial business associations improving the internationalisation of SMEs with digital platforms: A design science research approach. *International Journal of Information Management, 53*, 102070.
Cummings, J. L., & Holmberg, S. R. (2012). Best-fit alliance partners: the use of critical success factors in a comprehensive partner selection process. *Long Range Planning, 45*(2-3), 136-159.
Dawson, B. K., Young, L., Tu, C., & Chongyi, F. (2014). Co-innovation in networks of resources—A case study in the Chinese exhibition industry. *Industrial Marketing Management, 43*(3), 496-503.
de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: a research agenda. *Journal of Information Technology, 33*(2), 124-135.
Deeter-Schmelz, D. R., & Ramsey, R. P. (2003). An investigation of team information processing in service teams: Exploring the link between teams and customers. *Journal of the Academy of marketing Science, 31*(4), 409-424.
Deloitte. (2020). *COVID-19: Shaping the Future Through Digital Business*. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/global/Documents/About-Deloitte/COVID-19/gxtechnology-digital-COVID-19-recovery-shaping-future-digital-business.pdf, May 14, 2021.
Dwyer, F. R., Schurr, P. H., & Oh, S. (1987). Developing buyer-seller relationships. *Journal of marketing, 51*(2), 11-27.
Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of management review, 23*(4), 660-679.
Emden, Z., Calantone, R. J., & Droge, C. (2006). Collaborating for new product development: selecting the partner with maximum potential to create value. *Journal of product innovation management, 23*(4), 330-341.
Fonjong, B. B., Tian, H., & District, Z. (2019). Impact of value co-creation on firm performance: Mediating role of strategic advantages. *European Journal of Business and Management*, *11*(21),161-172.
Furlotti, M., & Soda, G. (2018). Fit for the task: Complementarity, asymmetry, and partner selection in alliances. *Organization science, 29*(5), 837-854.

Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of product innovation management, 31*(3), 417-433.

Goyal, S., Ahuja, M., & Kankanhalli, A. (2020). Does the source of external knowledge matter? Examining the role of customer co-creation and partner sourcing in knowledge creation and innovation. *Information & Management, 57*(6), 103325.

Hartwich, F., Engler-Palma, A., Gonzalez, C., Trujillo, L., & Bremmers, H. (2004). *Funding partnersmps for innovation: Did it bring about agroindustry development in Latin America.* Paper presented at the 6th International Conference on Chain and Network Management in Agribusiness and the Food Industry.

Hill, R. C., & Hellriegel, D. (1994). Critical contingencies in joint venture management: Some lessons from managers. *Organization science, 5*(4), 594-607.

Hillman, A. J., Withers, M. C., & Collins, B. J. (2009). Resource dependence theory: A review. *Journal of management, 35*(6), 1404-1427.

Hitt, M. A., Dacin, M. T., Levitas, E., Arregle, J.-L., & Borza, A. (2000). Partner selection in emerging and developed market contexts: Resource-based and organizational learning perspectives. *Academy of Management journal, 43*(3), 449-467.

Jayashankar, P., Nilakanta, S., Johnston, W. J., Gill, P., & Burres, R. (2018). IoT adoption in agriculture: the role of trust, perceived value and risk. *Journal of Business & Industrial Marketing*.

Jin, H., & Hurd, F. (2018). Exploring the impact of digital platforms on SME internationalization: New Zealand SMEs use of the Alibaba platform for Chinese market entry. *Journal of Asia-Pacific Business, 19*(2), 72-95.

Karhu, K., & Ritala, P. (2020). Slicing the cake without baking it: Opportunistic platform entry strategies in digital markets. *Long Range Planning*, 101988.

Kelley, S. W., Donnelly Jr, J. H., & Skinner, S. J. (1990). Customer participation in service production and delivery. *Journal of retailing, 66*(3), 315.

Kozarkiewicz, A., & Kabalska, A. (2016). Project Strategy and Strategic Project Management: The Understanding and the Perception of Relevance among Polish Project Management Practitioners. *Project Management Development–Practice and Perspectives*, 187.

Kozarkiewicz, A., & Kabalska, A. (2017). Kuracjusz i klient: ewolucja propozycji wartości w modelach biznesowych przedsiębiorstw uzdrowiskowych w Polsce. *Marketing i Rynek*(10 (CD)), 296-309.

KPMG. (2020). *Technology industry CEO outlook*. Retrieved from https://assets.kpmg/content/dam/kpmg/xx/pdf/2020/10/technology-industry-ceo-outlook.pdf, May 14, 2021.

Lee, S., Kim, B. G., & Kim, H. (2012). An integrated view of knowledge management for performance. *Journal of Knowledge management*, *33*(6), 804-821.

Li, S. X., & Rowley, T. J. (2002). Inertia and evaluation mechanisms in interorganizational partner selection: Syndicate formation among US investment banks. *Academy of Management journal, 45*(6), 1104-1119.

Lovelock, C., & Wright, L. (2001). *Principles of service marketing and management*: Prentice Hall.

Luo, Y. (2002). Contract, cooperation, and performance in international joint ventures. *Strategic Management Journal, 23*(10), 903-919.

Lusch, R. F., Vargo, S. L., & O'brien, M. (2007). Competing through service: Insights from service-dominant logic. *Journal of retailing, 83*(1), 5-18.

Maryono, M., Ghozali, I., Kusumawardhani, A., & Prabantariksa, M. (2019). How Value Network-Based Co-Inovation Contribute in Competitive Advantage. *JURNAL BISNIS STRATEGI, 28*(2), 79-90.

McKinsey. (2020). *The technology industry in 2020 and beyond*. Retrieved from https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/thetechnology-industry-in-2020-and-beyond#, May 14, 2021.
McKnight, D. H., Lankton, N. K., Nicolaou, A., & Price, J. (2017). Distinguishing the effects of B2B information quality, system quality, and service outcome quality on trust and distrust. *The Journal of Strategic Information Systems, 26*(2), 118-141.
McQuiston, D. H. (2001). A conceptual model for building and maintaining relationships between manufacturers' representatives and their principals. *Industrial Marketing Management, 30*(2), 165-181.
Moorman, C., & Rust, R. T. (1999). The role of marketing. *Journal of marketing, 63*(4_suppl1), 180-197.
Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management annals, 5*(1), 373-412.
Murphy, M., Perrot, F., & Rivera-Santos, M. (2012). New perspectives on learning and innovation in cross-sector collaborations. *Journal of Business Research, 65*(12), 1700-1709.
Parida, V., Westerberg, M., & Frishammar, J. (2012). Inbound open innovation activities in high-tech SMEs: the impact on innovation performance. *Journal of small business management, 50*(2), 283-309.
Parise, S., & Casher, A. (2003). Alliance portfolios: Designing and managing your network of business-partner relationships. *Academy of Management Perspectives, 17*(4), 25-39.
Parmigiani, A. (2007). Why do firms both make and buy? An investigation of concurrent sourcing. *Strategic Management Journal, 28*(3), 285-311.
Payne, A. F., Storbacka, K., & Frow, P. (2008). Managing the co-creation of value. *Journal of the Academy of marketing Science, 36*(1), 83-96.
Perry, A. G., Potter, P., A. , & Ostendorf, W. (2019). *Nursing Interventions & Clinical Skills E-Book*: Elsevier Health Sciences.
Powell, T. C., & Dent-Micallef, A. (1997). Information technology as competitive advantage: The role of human, business, and technology resources. *Strategic Management Journal, 18*(5), 375-405.
Queiroz, M., Tallon, P. P., Coltman, T., Sharma, R., & Reynolds, P. (2020). Aligning the IT portfolio with business strategy: Evidence for complementarity of corporate and business unit alignment. *The Journal of Strategic Information Systems, 29*(3), 101623.
Reid, S., Hayes, J., & Stibbe, D. (2015). Platforms for partnership: Emerging good practice to systematically engage business as a partner in development. *The Partnering Initiative*.
Ruggieri, R., Savastano, M., Scalingi, A., Bala, D., & D'Ascenzo, F. (2018). The impact of digital platforms on business models: An empirical investigation on innovative start-ups. *Management & Marketing, 13*(4).
Salancik, G. R., & Pfeffer, J. (1978). *The external control of organizations: A resource dependence perspective*: New York: Harper & Row.
Scheuing, E. E., & Johnson, E. M. (1989). A proposed model for new service development. *Journal of Services Marketing,3*(2).
Scott-Kennel, J. (2013). Models of internationalisation: the New Zealand experience. *International Journal of Business and Globalisation, 10*(2), 105-136.
Selnes, F., & Gønhaug, K. (2000). Effects of supplier reliability and benevolence in business marketing. *Journal of Business Research, 49*(3), 259-271.
Shah, R. H., & Swaminathan, V. (2008). Factors influencing partner selection in strategic alliances: The moderating role of alliance context. *Strategic Management Journal, 29*(5), 471-494.

Snow, C. C., Miles, R. E., & Miles, G. (2005). A configurational approach to the integration of strategy and organization research. In: Sage Publications London, Thousand Oaks, CA and New Delhi.

Spin, M. (2012). *The effect of intra-organizational routines and inter-organizational routines on collaborative innovation performance. An experiment.* University of Twente,

Subramani, M. R., & Venkatraman, N. (2003). Safeguarding investments in asymmetric interorganizational relationships: Theory and evidence. *Academy of Management journal, 46*(1), 46-62.

Tian, J., Wang, K., Chen, Y., & Johansson, B. (2010). From IT deployment capabilities to competitive advantage: An exploratory study in China. *Information Systems Frontiers, 12*(3), 239-255.

Tippins, M. J., & Sohi, R. S. (2003). IT competency and firm performance: is organizational learning a missing link? *Strategic Management Journal, 24*(8), 745-761.

Tsou, H.-T., Cheng, C. C., & Hsu, H.-Y. (2015). Selecting business partner for service delivery co-innovation and competitive advantage. *Management Decision*.

Van de Vrande, V., De Jong, J. P., Vanhaverbeke, W., & De Rochemont, M. (2009). Open innovation in SMEs: Trends, motives and management challenges. *Technovation, 29*(6-7), 423-437.

Wade, M., & Hulland, J. (2004). The resource-based view and information systems research: Review, extension, and suggestions for future research. *MIS quarterly*, 107-142.

Wang, H. C., He, J., & Mahoney, J. T. (2009). Firm-specific knowledge resources and competitive advantage: the roles of economic-and relationship-based employee governance mechanisms. *Strategic Management Journal, 30*(12), 1265-1285.

Wang, L., & Zajac, E. J. (2007). Alliance or acquisition? A dyadic perspective on interfirm resource combinations. *Strategic Management Journal, 28*(13), 1291-1317.

Weber, J. A. (2001). Partnering with resellers in business markets. *Industrial Marketing Management, 30*(2), 87-99.

Yeniyurt, S., Henke, J. W., & Yalcinkaya, G. (2014). A longitudinal analysis of supplier involvement in buyers' new product development: working relations, inter-dependence, co-innovation, and performance outcomes. *Journal of the Academy of marketing Science, 42*(3), 291-308.

Zeithaml, V. A., Parasuraman, A., & Malhotra, A. (2002). Service quality delivery through web sites: a critical review of extant knowledge. *Journal of the Academy of marketing Science, 30*(4), 362-375.

DECISION SCIENCES INSTITUTE

Knee Optimization for Queuing Systems: A Customized Approach

Danqi Lu
Boston University Metropolitan College
Email: danqilu@bu.edu

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

Canan Gunes Corlu
Boston University Metropolitan College
Email: canan@bu.edu

ABSTRACT

An empirical study is presented for finding the “knee” of a queueing system’s “hockey-stick” curve that describes the relationship between server utilization and customer waiting time. Several mathematical approaches for finding the knee are compared to knee values chosen by human subjects for a range of scenarios. The most precise approach is based on the Euclidean distance from the curve’s fitted regression line to the farthest point on the curve. This method can be calibrated based on inputs by managers of a targeted service organization and used in a decision support system for capacity planning.

KEYWORDS: Queuing, Knee, Optimization, Waiting, Capacity

INTRODUCTION

Queuing systems have been studied by researchers for over 100 years (Asmussen & Boxma, 2009). Yet, excessive waiting remains among the most frequent customer service complaints in call centers (Sarel & Marmorstein, 1998), hospitals (Bleustein et al, 2014), and government services (Liu & Yen, 2016), among others. Secondary impacts of excessive customer waiting include decreases in customer perceptions of service quality (Tiwari et al, 2017), decreases in customer satisfaction (Garcia et al, 2012; Peevers et al, 2009; Davis & Volmann, 1990), negative impacts on customer loyalty (Bielen & Demoulin, 2007), and increased stress among service workers (Spigg & Jackson, 2006). One reason for long waiting times is the inability of service managers to reconcile uncertainties while making capacity planning decisions (Labach, 2010). Their task is made especially challenging because of the prevalence of performance metrics that primarily measure server productivity (Tuten & Neidermeyer, 2004).

The primary focus of queuing system optimization is minimizing a function that includes the cost to provide a service and the cost associated with customer’s waiting or turnaround times (i.e., from arrival to completion of service). Although the mathematical expression of these decisions is relatively straight-forward, the cost associated with customer’s perception of waiting can be difficult if not impossible to quantify (Kembe et al, 2012; Fenton et al, 2012). The challenge for service managers is exacerbated by the non-linear hockey-stick curve prevalent in queuing systems (Maleyeff, 2021, p. 131). The point on this curve where it most noticeably bends is referred to as the knee of the hockey-stick curve (this point has also been called the elbow). It is the location often recommended as the target server utilization. The need to identify the knee of similar curves is also present in other decision settings, including those associated with clustering (Salvador & Chan, 2004), internet broadband (Kleinrock, 2018), and image processing (Bhandari et al, 2016).

In this article, we use an empirical approach to determine the most precise mathematical method for determining the knee of a queuing system hockey-stick curve, based on a comparison of choices made by a set of human subjects. The subjects are practitioners who appreciate the challenge of optimizing resource productivity while providing effective service to customers. The results are used to create a flexible but customized decision support application for capacity planning. This article is organized as follows. First, queuing systems are described and their analysis approaches are reviewed, with emphasis on the hockey-stick curve. Second, a collection of mathematical approaches for finding the knee are described. Third, our methodology is presented, including the survey mechanism and the empirical analysis approach. Fourth, results are analyzed and the most precise mathematical approach for finding the knee is identified. Fifth, customization of the mathematical approach is described, followed by an example of how our results could be used to inform a decision support system for capacity planning.

BACKGROUND

Queuing systems are characterized by their structure and the uncertainty associated with their operation. They range from single-server, single-stage systems to multiple-server, multi-echelon systems. The uncertainty associated with a queuing system mainly concerns the arrival patterns of customers, service time variation, and the potential for queue abandonment by customers. A queuing system includes a discipline that dictates the order with which customers are served. Customers who exit the system may at times be routed back into the system. Other structural elements include the potential batching of arrivals, the size of the customer population, and the number of customers that the queuing system can hold.

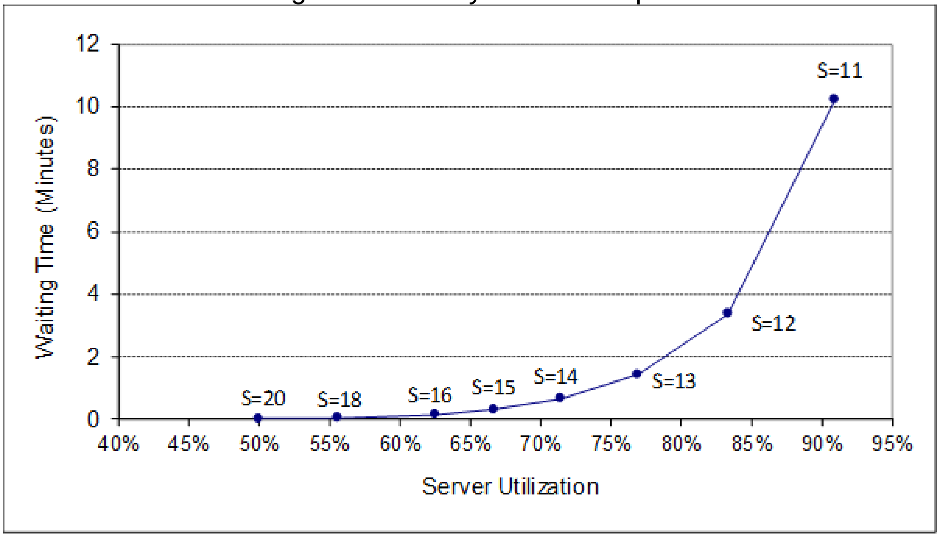
Queuing systems are analyzed using one of two alternative approaches. The first approach uses a closed-form mathematical expression that is easy to apply but only useful when the system under study adheres to the model's strict assumptions. They can also provide approximate solutions in cases where precision is not required. The second approach uses a more robust Monte Carlo simulation (MCS). These queuing systems may include multiple stages, unique customer arrival patterns, and statistical variations that do not adhere to the limited assumptions present in mathematical queuing models. Because they transform random inputs to performance characteristics, the random variation of simulation output needs to be taken into consideration by the analyst.

Coverage of the M/M/S analytical queuing model is a mainstay of undergraduate and graduate operations management courses. This model assumes that customers arrive according to a Poisson process, they wait in a single queue for one of S servers, and the service time follows an exponential distribution. The model is valid for any queue discipline. It assumes that the customer population is infinite and that the system has no queue size limitations. Like all approaches to queuing system analysis, the model predicts statistical parameters associated with queue length, customer waiting times, customer time spent in the system, among other specialized performance criteria.

The two main perspectives that need to be reconciled when implementing a decision support system (DSS) for a queueing system are the service provider perspective (the desire to minimize resource costs) and the customer perspective (the desire for prompt service). Service provider costs are often well-known, consisting of labor or equipment costs. A useful operational performance measure corresponding to this cost is server utilization (i.e., the percentage of time that servers are interacting with customers). The cost of waiting is usually difficult or impossible to quantify, and will vary by application. Rather than rely on these cost estimates, the method described below indirectly accounts for them when determining the optimal server utilization. The main operational measures that correspond to the cost of waiting are the queue length characteristics and the customer waiting time characteristics (e.g., average, percentiles, etc.).

The non-linear hockey-stick relationship of server utilization and the average customer waiting time is present in every queuing system with uncertainty in arrivals and/or service times. Consider a service where 40 customers arrive per hour and each server has the capacity to serve 4 customers per hour. Application of the M/M/S model provided the results shown in Figure 1, where the number of servers (S) is changed from 20 to 11. Presented with this hockey-stick curve, a manager would be able to choose the number of servers that they believe effectively balances cost and service. For example, a manager with less concern for prompt service or one with high labor costs may choose 12 servers. Alternatively, a manager with more concern for prompt service or one with low labor costs may choose 13 servers.

Figure 1: Hockey-Stick Example



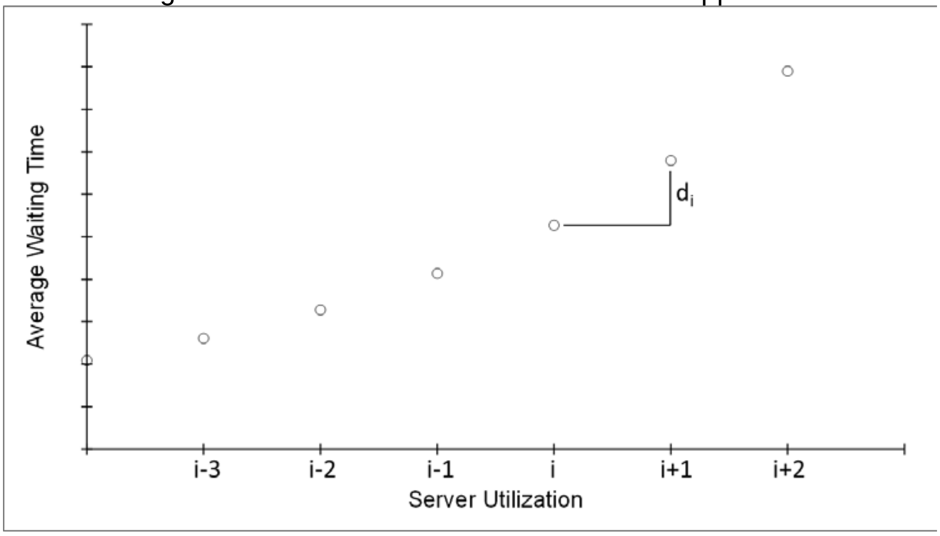
KNEE DETERMINATION METHODS

A human possesses the cognitive pattern recognition skills to identify the knee of a curve using a consistent approach. Applying a mathematical approach to find the knee usually involves searching a point representing the most abrupt change on the slope of the curve. The mathematical methods described below are used to find the knee for a finite set of points that follow a hockey-stick pattern.

Approaches Based on Derivatives

Because of its relationship to the rate of change of the curve's slope, many methods identify the knee based on the curve's first and/or second derivative at each potential knee value. Chiu et al. (2001) suggest that the knee can be found at the point with the largest magnitude between two adjacent points. As shown in Figure 2, this approach is simply the maximum value of d_i-d_{i-1} (where d_i is the positive vertical distance from point i to point i+1). Harris et al. (2000) suggests that the knee be found based on the largest second derivative or, in the discrete case, the ratio of the distance between adjacent points (in this case the ratio of d_i and d_{i-1}). A related third approach is based on the lowest point with a second derivative above a specified threshold value (Foss & Zaïane, 2002).

Figure 2 : Framework for Derivative-Based Approach



Approaches Based on Linear Regression

Satopaa et al. (2011) suggest that the knee is the point that has the largest Euclidean distance from a linear equation that connects the first and last data point, as shown in Figure 3. A similar approach (Figure 4) was suggested by Salvador and Chen (2004) whereby a linear regression equation is fit to the entire curve, with the knee located at the point representing the largest Euclidean distance from the curve to the fitted line. For each of these methods, the setting of the leftmost and rightmost x-axis server utilizations could affect the knee value. Figures 3 and 4 fit the lines using data for server utilizations from 60% to 99%, in increments of 1%.

Salvador & Chen (2004) recommend an alternative approach called the L-method. This approach evaluates every potential pair of two contiguous sub-segments of the curve. For each pair of sub-segments, two least-squares linear regression equations are fit (one for the left sub-segment and one for the right sub-segment). The knee is identified based on the iteration with the lowest combined root mean squared error. Figure 5 illustrates this method by showing one iteration.

Figure 3: Approach Based on Endpoint Connection

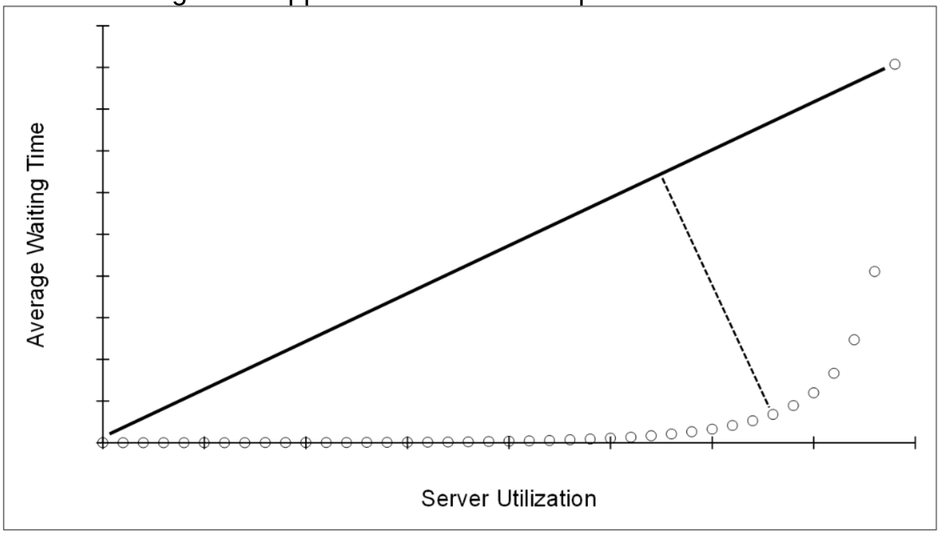
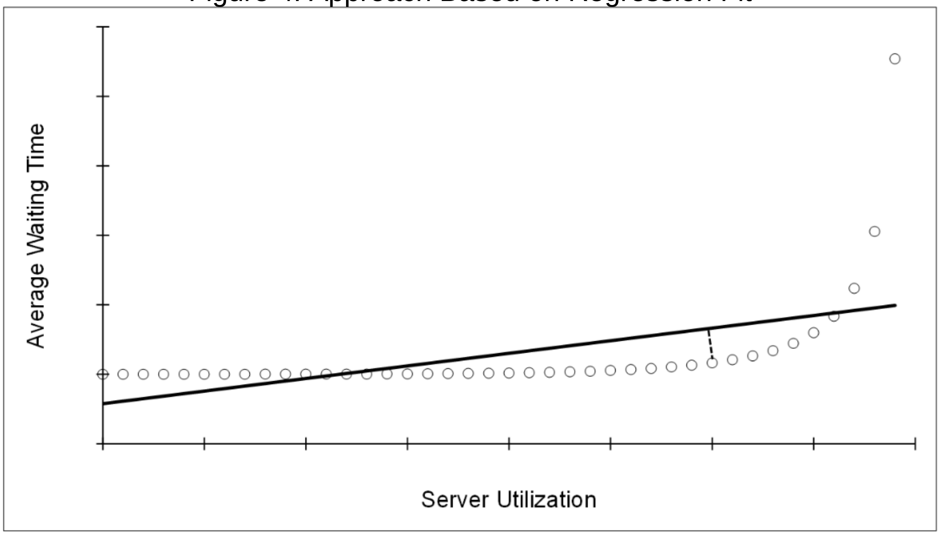


Figure 4: Approach Based on Regression Fit

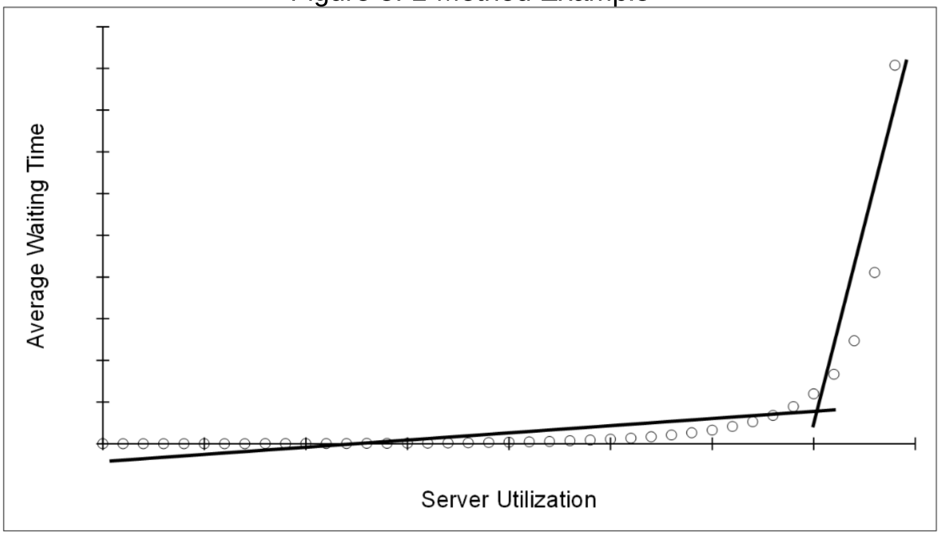


Approach Based on Power Function

The last approach was recommended by Kleinrock (2018) for capacity determination to prevent Internet congestion. Here, a power function is calculated for each potential server utilization. As shown in Equation 1, W_S is the average customer time in system (i.e., turnaround time), μ is the average server capacity (e.g., customers/hour), and ρ is the server utilization. The numerator corresponds to server cost and will vary from 0 to 1. It represents a criterion that we wish to maximize (Kleinrock referred to this as “good”). The denominator creates a normalized measure of service promptness that will vary from 1 to about 2. It represents a criterion that we wish to minimize (Kleinrock referred to this as “bad”). Thus, the Power function is the ratio of good to bad, and this ratio would be maximized at the knee.

$$Power = \frac{\rho}{\mu W_S} \quad (1)$$

Figure 5: L-Method Example

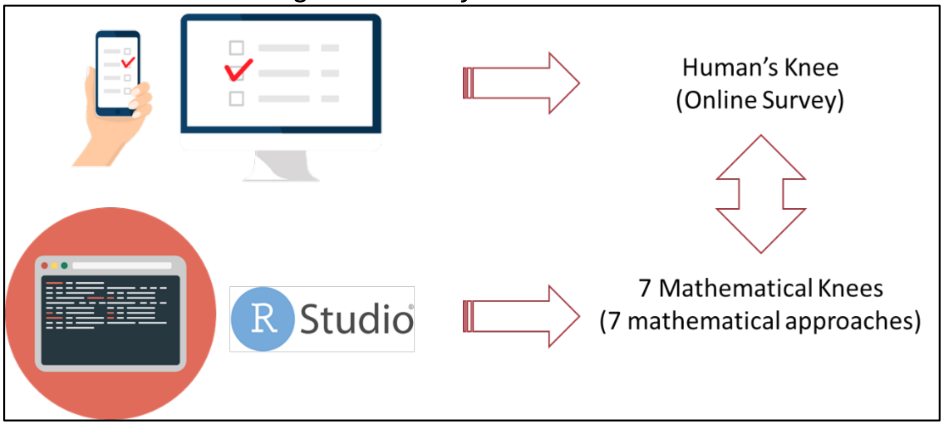


The challenge of implementing a mathematical approach to finding the optimal server utilization is that capacity planning decision making can encounter widely different cost structures. Any mathematical approach that does not take costs into account (directly or indirectly) will suffer from inaccuracy. Consider staffing a reception desk with interns at an investment company's walk-in financial services office. Customers will be investors (many with high net worth) while the interns are low paid service providers. The manager of the financial center would be well served by allocating many interns to the reception desk even though their labor utilization may be low. The financial center can be contrasted with a call-in legal service with very high labor costs and customers without expectations of especially fast service. In this case, the best planned server utilization will likely be high in order to minimize labor costs. The method proposed below will maintain its mathematical orientation while being customizable for specific applications.

METHODOLOGY

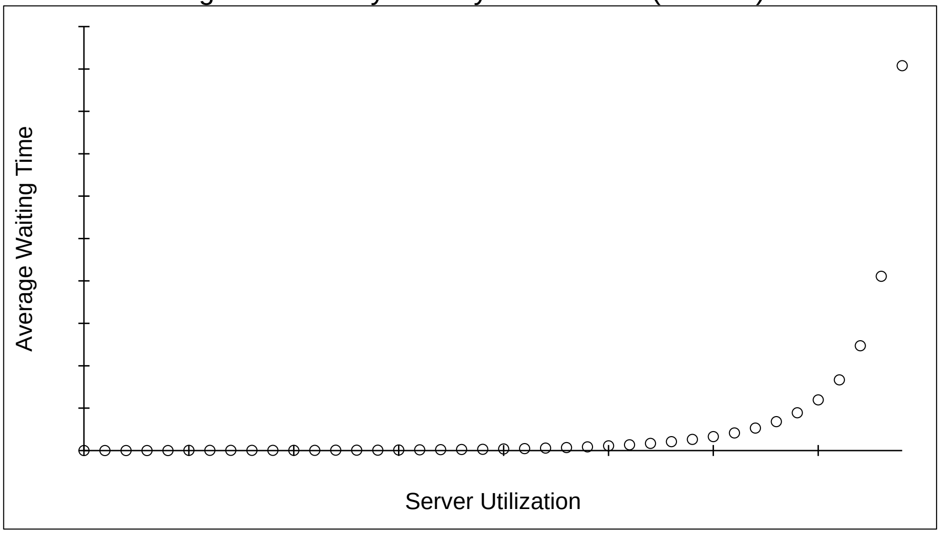
The framework for our analysis approach is described in Figure 6. It consists of finding the mathematical knee method that most precisely matches the choices made by a set of working professionals who participated in an online survey. The survey began with an explanation of the meaning of a hockey-stick curve as applied to a queuing system and a video describing how to take the survey. Each subject was asked to choose the optimal server utilization for a set of 32 curves, representing 16 different scenarios, each repeated twice. The online survey application was developed using R Shiny software.

Figure 6: Analysis Framework



The curves all resembled Figure 7, with no scales indicated in order to minimize bias associated with participants' individual experiences as service providers, managers, or customers. Participants clicked on the point they would consider as the optimal server utilization, before moving to the next curve. Their 32 choices for optimal server utilization were downloaded for later analysis.

Figure 7: Survey Hockey-Stick Curve (M/M/50)



Of the 16 different hockey-stick curves presented to survey participants, four were derived from M/M/S queuing models (with S=1, S=5, S=10, and S=50). The curve in Figure 7, for example, was derived using a M/M/50 model. Only four M/M/S hockey-stick curves were shown because M/M/S curves are remarkably similar when axes labels are removed. The remaining 12 curves were created with varying degrees of curve bending. Because these 12 curves were not actual queuing hockey-stick curves, they serve the purpose of eliminating bias based on the potential for subjects' remembering previous choices as they progress through the survey.

The knee values chosen by the survey participants were compared to the mathematical knee methods described above and shown in Table 1. Methods M4 and M5 used the data starting from server utilizations of 50% to 99%. The analysis of the results focused on the method that most consistently matched the human choices for the knee, by determining the accuracy and precision of each method. Accuracy (i.e., the average difference between the method and the subjects' choice) is less important than precision (i.e., the variation between the method and the subjects' choice). As with other measurement systems, the method will be calibrated for specific applications and, therefore, precision is the most important criterion for determining the best mathematical approach.

Table 1: Mathematical Knee Methods

Method	Description
M1	Largest magnitude slopes of adjacent points (difference of first derivatives)
M2	Largest ratio slopes of adjacent points (ratio of first derivatives)
M3	Largest ratio difference of adjacent points (ratio of second derivatives)
M4	Largest Euclidean distance from linear fit of first & last fitted point
M5	Largest Euclidean distance from linear fit of all points
M6	L-method
M7	Maximum power function

RESULTS

Fifty-four participants completed the survey. Of these, four participants provided data that clearly showed misunderstanding of the instructions; they were removed from the analysis. The remaining 50 participants are analyzed below. Figure 8 shows a typical result for one survey participant. Because each of the curves included in the display was shown to each participant twice, two knee results are shown (when one point is shown for a curve number, the knee choices were identical). The set of curves included in Figure 8 were not the M/M/S curves; they were designed to increase in hockey-stick curvature from almost a constant tangent to almost an L shape. The expectation was that the variation of pairs of choices would increase and the knee would increase from the first to the tenth curve. These curves served the dual purpose of eliminating bias and confirming that the results trended as expected.

Table 2 shows the key numerical results for the four M/M/S hockey-stick curves. All of the methods showed some level of bias, as evidenced by the average difference between the human subjects' choices and the mathematical method. As anticipated above, this level of bias is not concerning. The precision of each method is critical, however. Precision is measured by the standard deviation of the differences between the subjects' choice and the mathematical calculation. The results indicate that methods M4 and M5 are the more precise and that they are essentially equivalent in precision. Although they appear equivalent in precision, Method M4 is biased low, while Method M5 is biased high.

Because hockey-stick curves can also be generated by a MCS, Method M5 is the most practical method for finding the knee. The rationale for this choice is as follow. When simulating, each predicted average waiting time result is subject to random variation. Importantly, the amount of variation is much higher for scenarios with high server utilizations (Corlu et al, 2021). For simulated results, the linear regression line for method M5 (that uses all of the hockey-stick curve data) will be preferred to the line generated by method M4 (that uses just the first and last point on the curve). Method M4 would be especially problematic because the rightmost average waiting time will be impacted by large random variations that will affect the accuracy of a line connecting just the first and last points on the curve.

Figure 8: Example Survey Results

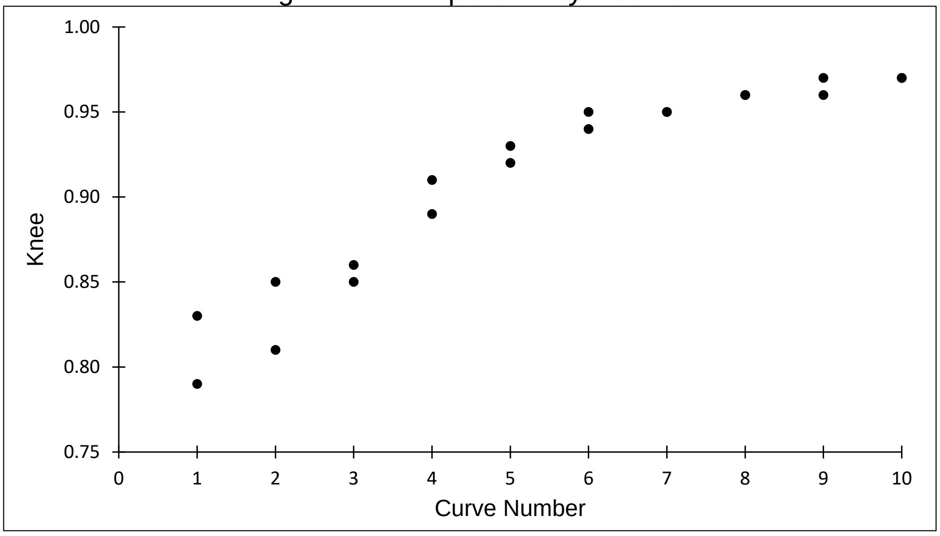


Table 2: Results of Comparing Human Choices to Mathematical Methods

Method	Knee Values				Distances from Calculated Knee	
	M/M/1	M/M/5	M/M/10	M/M/50	Average	Std Dev
M1	0.99	0.99	0.99	0.99	-0.0795	0.0086
M2	0.98	0.98	0.98	0.98	-0.0695	0.0086
M3	0.97	0.97	0.97	0.97	-0.0595	0.0086
M4	0.93	0.93	0.93	0.94	-0.0220	0.0038
M5	0.87	0.87	0.87	0.88	0.0380	0.0038
M6	0.94	0.94	0.94	0.94	-0.0295	0.0086
M7	0.50	0.69	0.76	0.89	0.2005	0.1563
Participant Average	**0.906**	**0.904**	**0.908**	**0.923**		

DSS APPLICATION

The mathematical knee method will be embedded in a customized DSS for capacity planning. The model would generate the optimal server utilization based on calibration by the user. The procedure for calibrating the knee calculation would be consistent with the survey approach used above. The manager of the service for which the DSS is designed would choose the optimal server utilization for each of a set of hockey-stick curves derived from modeling their queuing system. These curves would include scales to show the manager the server utilizations and customer waiting times corresponding to their operation The manager's choices would represent their most effective reconciliation of server cost (i.e., server utilization) and customer promptness (i.e., average waiting time). As shown in Figure 9, a manager with lower server costs and/or a desire for prompt service would tend to choose the responsiveness preference, while managers with high server costs and/or less of a need for promptness of service would tend to choose the cost efficiency preference.

Figure 9: Comparison of Service Managers' Calibration

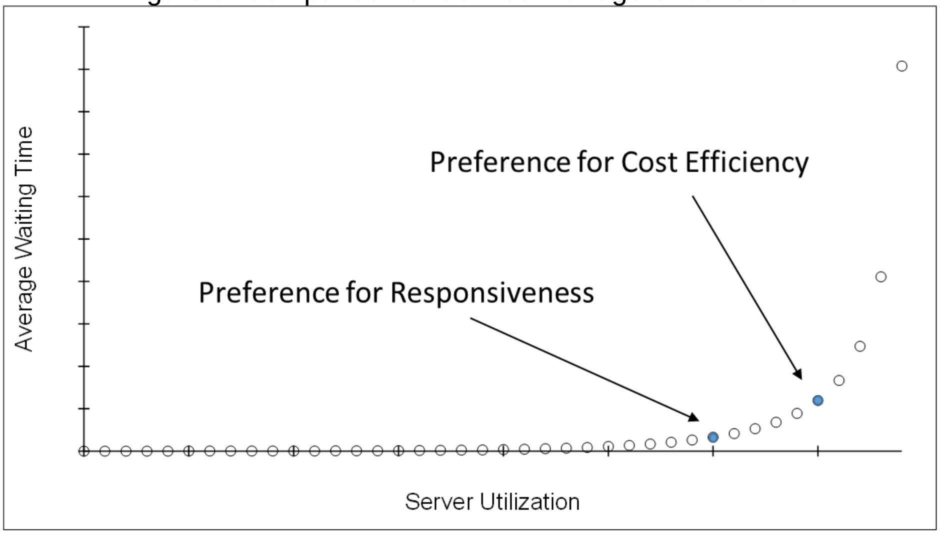


Figure 10 shows an example of one iteration of the calibration procedure. In this case, the hockey-stick curve was derived based on a MCS. It represents one potential scenario based on an assumed arrival time, service rate, and other uncertain parameters. The linear regression fit (based on method M5) is given in Equation 2 (where Wq is the average waiting time and ρ is the server utilization). The longest Euclidean distance from the line to the curve is 0.0749, which occurs at server utilization 80% (this is the mathematical knee). In the example, the manager chose a utilization of 84% as their preferred target (i.e., the red dot). If the average differential (mathematical knee vs. manager choice) of all iterations is the same 4%, the DSS would recommend an optimal server utilization based on the mathematical knee, with a positive adjustment of 4%.

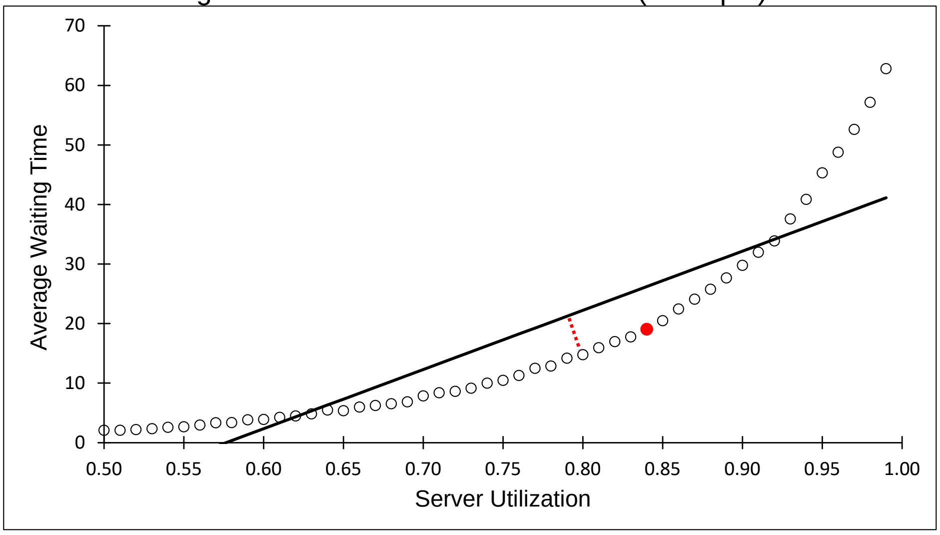
$$W_q = 99.4045\rho - 57.299 \tag{2}$$

CONCLUSION

This work concerned the development of a customized method for determining the optimal server utilization based on the mathematical knee of a queuing model hockey-stick curve. An empirical experiment compared seven mathematical approaches with choices made by human survey participants. The preferred method has the best precision when compared to human choices. It is both automated and customized when using a calibration procedure. The method would be implemented into a DSS so that a manager would optimize labor allocation using a consistent solution criterion.

Future work will concentrate on standardizing and integrating the calibration procedure so that it can be embedded automatically into a capacity planning DSS. For example, the system may include a calibration survey whereby results are automatically translated to the optimal server utilization. The system used by managers to plan capacity could be optimized when conditions warrant, or it can be setup as interactive with each user-entered allocation plan scored against optimal knee values. The latter option would be preferred by managers who seek to take nuances into account, while at the same time having a system that evaluates their plan and offers suggestions for improved resource allocation.

Figure 10: Customization Procedure (Example)



REFERENCES

Asmussen, S., & Boxma, O.J. (2009). Editorial introduction. *Queueing Systems 63*(1-4). 1-2.

Bleustein, C., Rothschild, D.B, Valen, A., Valatis, E., Schweitzer, L., & Jones, R. (2014). Wait times, patient satisfaction scores, and the perception of care. *The American Journal of Managed Care, 20*(5), 393–400.

Bhandari, A.K., Kumar, A., Singh, G.K., & Soni, V. (2016). Dark satellite image enhancement using knee transfer function and gamma correction based on DWT–SVD. *Multidimensional Systems and Signal Processing, 27*(2), 453–476.

Bielen, F., & Demoulin, N. (2007). Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality: An International Journal*, *17*(2), 174-193.

Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001). A robust and scalable clustering algorithm for mixed type attributes in large database environment. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 263–268.

Corlu, C.G., Maleyeff, J., Yang, C., Ma, T., & Shen, Y. (2021). Decision support system with simulation-based optimization for healthcare capacity planning. Proceedings of the 10th Simulation Workshop of the Operational Research Society, Leicestershire, UK.

Davis, M. M., & Vollmann, T. E. (1990). A framework for relating waiting time and customer satisfaction in a service operation. *The Journal of Services Marketing*, *4*(1), 61.

Fenton, J.J., Jerant, A.F., Bertakis, K.D., & Franks, P. (2012). The Cost of Satisfaction. *Archives of Internal Medicine, 172*(5), 405-411.

Foss A., & Zaïane, A. (2002). A parameterless method for efficiently discovering clusters of arbitrary shape in large datasets. IEEE International Conference on Data Mining Proceedings, 179-186.

Garcia, D., Archer, T., Moradi, S., & Ghiabi, B. (2012). Waiting in vain: Managing time and customer satisfaction at call centers. *Psychology*, *3*(2), 213-216.

Harris, R. S, Hess, D. R, & Venegas, J. G. (2000). An objective analysis of the pressure-volume curve in the acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine, 161*(2), 432–439.

Kembe, M. M., Onah, E. S., & Iorkegh, S. (2012). A study of waiting and service costs of a multi-server queuing model in a specialist hospital. *International Journal of Scientific & Technology Research, 1*(8), 19-23.

Kleinrock, L. (2018). Internet congestion control using the power metric: Keep the pipe just full, but no fuller. *Ad Hoc Networks*, *80*, 142-157.

Labach, E. (2010). Improving customer retention through service quality at call centers. *International Journal of Management and Information Systems*, *14*(3), 71-76.

Liu, W. & Yen, C. (2016). Optimizing bus passenger complaint service through big data analysis: Systematized analysis for improved public sector management. *Sustainability (Basel, Switzerland)*, *8*(12), 1319.

Maleyeff, J. (2021). Ser*vice Science: The Analysis and Improvement of Business Processes*, UK: Routledge.

Peevers, G., McInnes, F., Morton, H., Matthews, A., & Jack, M. (2009). The mediating effects of brand music and waiting time updates on customers' satisfaction with a telephone service when put on-hold. *The International Journal of Bank Marketing*, *27*(3), 202-217.

Salvador, S., & Chan, P. (2004). Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. 16th IEEE international conference on tools with artificial intelligence, 576-584.

Sarel, D., & Marmorstein, H. (1999). Managing the delayed service encounter: The role of employee action and customer prior experience. *The International Journal of Bank Marketing*, *17*(6), 286-294.

Satopaa, V., Albrecht, J., Irwin, D., & Raghavan, B. (2011). Finding a “kneedle” in a haystack: Detecting knee points in system behavior. 31st International Conference on Distributed Computing Systems Workshops, 166-171.

Sprigg, C.A., & Jackson, P.R. (2006). Call centers as lean service environments: Job-related strain and the mediating role of work design, *Journal of Occupational Health Psychology*, 11(2), 197-202.

Tiwari, V., Queenan, C., & St. Jacques, P. (2017). Impact of waiting and provider behavior on surgical outpatients’ perception of care. *Perioperative Care and Operating Room Management*, *7*, 7–11.

Tuten, T., & Neidermeyer, P. (2004). Performance, satisfaction and turnover in call centers: The effects of stress and optimism. *Journal of Business Research*, *57*(1), 26-34.

DECISION SCIENCES INSTITUTE

The Performance Implication of The Form of Information Technology Organization (ITO): A Resource-Based Perspective

Yu Liu
Saginaw Valley State University
Email: yliu1238@svsu.edu

ABSTRACT

In this study, we view the form of IT organization as a platform for developing complementary IT resources and capabilities. By using the resource-based view, we argue that the approach to organizing IT activities may have a positive effect on firm performance under certain conditions. We tested the hypotheses on a sample of 798 observations and found that a more centralized form of IT organization leads to positive performance effects when the scale of IT investment is high. However, a more decentralized IT organization leads to positive performance effects when the level of firm diversification is high.

KEYWORDS: IT Organization, Resource-based View, IT Investment, Firm Diversification, Firm Performance

INTRODUCTION

Prior information systems (IS) research has shown the importance of running an effective information technology (IT) organization to meet a firm's strategic and organizational needs. A strong focus in the previous studies has been on the mode of IT governance. Yet a salient research gap in the literature and hence a promising opportunity to pursue is how a firm's choice of organizing IT activities interacts with other organizational factors to affect firm performance. There has been limited research that has examined the performance impact of IT governance modes, let alone the possible interactions between IT governance and certain organizational attributes. In the meantime, arguments related to the choice of IT governance modes in the past have often referred to the concept of fit, which has apparent performance implications. In other words, the concept of fit suggests that the effectiveness of a certain IT governance mode will depend on organizational contingencies such as those represented by a firm's IT and business strategies (Brown & Grant, 2005).

As an attempt to fill the research void, the authors study how the organizational arrangement of IT activities affects firm performance via the interaction with other variables such as the scale of IT investment and firm diversification. Past research has shown that the mode of IT governance is important for realizing business value from a firm's IT investment (Weill, 2004). Recent studies have added more to our understanding of how the form of IT governance affects organizational performance. For example, Turel et al. (2017) found that board-level IT governance influenced organizational performance through the mediation of strategic alignment. Moreover, Turedi and Zhu (2019) found out that IT investment interacted with IT governance to create value for the firm. A large-scale IT investment is often made in an attempt to integrate the various functions and business units of a firm. Given the high level of such investments, it is of critical importance for centralizing a firm's IT capabilities to effectively coordinate such efforts. By contrast, a decentralized arrangement of IT activities may not be able to reap the benefits of enterprise-wide large-scale IT investments.

Meanwhile, the organizational arrangement of IT activities may also interact with a firm's diversification strategy. Firm diversification has been studied in the previous IT literature as an important strategic variable that intertwines with IT variables such as IT investments (Dewan et al., 1998), IT business value (Ravichandran et al., 2009), and the like. Previous research has suggested that IT investments need to be adjusted or fine-tuned to meet the need of diversification, particularly the need to use certain organizational control systems to manage the diversified business portfolio (Ravichandran et al., 2009). By the same token, the arrangement of IT activities may also need to be adjusted to accommodate the needs of different types and levels of diversification. Such an adjustment may also be viewed as a form of strategic alignment of IT or business-IT alignment. The existing literature on strategic alignment of IT has shown that despite being a challenge in practice, the strategic alignment between business and IT often has a positive effect on organizational performance (Sabherwal et al., 2019).

By using the resource-based view (RBV) and the concept of fit, this study investigates the impact of the organizational arrangement of IT activities on firm performance. In particular the authors adopt a contingency perspective and argue that the organizational form of IT activities would interact with firm characteristics such as IT investment and firm diversification to affect firm performance, perhaps by developing the needed IT resources and capabilities for better business-IT alignment. As one among the few to unravel the performance implications of IT governance choices, it is hoped that this study would motivate future researchers to explore the matter in a much deeper manner.

THEORY AND HYPOTHESES

Organizing IT Activities for Complementary IT Capabilities

A traditional view of the organizational arrangement of IT activities has been focused on the IT governance mode, differentiated by the locus of authority for IT activities (Sambamurthy & Zmud, 2000). Often defined as centralized, decentralized, and federal, the mode of IT governance has been found related to a variety of organizational factors such as organizational structures, business strategy, primary industry, firm size, and environmental conditions (Brown & Grant, 2005). A research commentary has called for a new organizing logic for IT activities, i.e. the platform logic (Sambamurthy & Zmud, 2000). By viewing the organizational arrangement of IT activities as a platform for developing important IT capabilities, the authors suggest that the organizing logic of IT activities should lead to those IT capabilities that are vital for business success through deploying IT initiatives. In a more recent study by Gregory et al. (2018), the authors argued that IT consumerization, a concept that captures the wide use of consumer devices and applications, might lead to IT governance transformation from functional IT governance to platform-based governance. Their study seems to suggest that firms are transitioning to the platform logic of IT activities given the ubiquitous existence of digital technologies these days.

Besides the transformation of IT governance forms, increasingly IS researchers have become interested in the business value of IT governance mechanisms. As a result, recently IT governance researchers have started to examine the link between IT governance forms and firm performance. For example, Wu et al. (2015) found that the effect of IT governance on firm performance was mediated by strategic IT alignment. In their event study of Asian firms, Chong and Duong (2017) found that while IT governance structural mechanism was associated with firm performance positively, IT governance process and relational mechanisms had a negative association with firm performance. Meanwhile, another group of researchers have looked at the relationship between inter-firm IT governance on various performance outcomes (Chi et al., 2017; Huber et al., 2017; Park et al., 2017). Despite these studies, this line of research has not produced consistent results regarding the performance implication of IT governance. Hence more research on the mechanism through which IT governance affects firm performance is warranted.

For years the IS literature has used the resource-based model to view the different IT resources and capabilities that may affect the overall firm performance (Melville et al., 2004). According to RBV capabilities are a firm's skills of coordinating and using resources in a productive manner in order to accomplish certain activities (Barney, 1991; Grant, 1991; Wernerfelt, 1984). By adopting this perspective, the organizational arrangement of IT activities can be seen as a managerial choice that interacts with other firm characteristics to develop complementary IT resources and capabilities for the firm. In this study the authors are interested in characteristics that may entail either a coordination need or a localization need. Particularly the authors investigate two factors that have seemed to be under-researched in the literature within such a context: the scale of IT investment and the degree of firm diversification. Although IT investment intensity has been shown to be strongly correlated with a firm's IT governance mechanisms (Ali & Green, 2012), an interactive effect of IT investment intensity and IT governance forms on firm performance has not been adequately addressed. Prior studies have shown that IT investment often interacts with other firm characteristics and their interactions are often found to impact firm performance (Ravichandran et al., 2009; Sabherwal et al., 2019). The effect of the interaction between IT investment and IT governance forms on firm performance thus presents an intriguing topic that warrants further investigation.

In large multi-business organizations, a centralized corporate-level IS unit often coexists with decentralized business-level IS units (Reynolds & Yetton, 2015). The organizational arrangement of IT activities and the allocation of IS human capital across an organizational hierarchy may manifest the management's effort to develop IT capabilities to meet different managerial needs. Centralized corporate-level IS units aim at realizing enterprise-wide economies and efficiencies. Therefore, a large centralized IS unit provide an important coordination capability across the organization. In the past, it has been shown that decisions regarding IT infrastructure development are often made by a central IS unit, which builds up the organization's coordination capabilities (Xue et al., 2011). In the meantime, decentralized IS units at the business unit level may supply the necessary IT capabilities for developing applications that address local business needs. Employees in a decentralized IS unit can work effectively with business-level managers and employees to create applications and tools to increase local responsiveness.

The interaction between the different arrangements of IS units and the scope of the firm's businesses can be viewed through the lens of strategic alignment of IT or business-IT alignment. Over the last nearly three decades since the seminal paper by Henderson and Venkatraman (1993), IS researchers have established an enduring theme of examining the antecedents and consequences of business-IT alignment. However, despite decades of research on strategic alignment of IT, the effect between strategic alignment of IT and firm performance has remained inconclusive (Queiroz, 2017). For example, in their study on business-IT alignment and organizational agility, Liang et al. (2017) found that different dimensions of IT alignment actually influenced a firm's agility differently. Meanwhile, many of the researchers have chosen to investigate the interactive effect of IT governance and strategic IT alignment. For example, Wu et al. (2015) showed that strategic IT alignment could mediate the effectiveness of IT governance on firm performance. More recently Héroux & Fortin (2018) examined the moderating role of IT-business alignment in the relationship between IT governance, IT competence, and innovation. Another study by Raymond et al. (2019) surveyed 223 manufacturing small and medium enterprises and examined the relationship between a firm's IT governance, IT alignment capabilities, and its IT performance.

Scale of IT Investment

The scale of IT investment is an important dimension of a firm's IT investment strategy (Ravichandran & Liu, 2011). Recent IS research has started to look at a firm's digital or IT strategic posture (Xue et al., 2017), of which IT investment is an important component. According to Ravichandran and Liu (2011), a large-scale IT investment usually indicates a proactive strategic posture towards IT. The scale or intensity of IT investment has also been the focus of a large body of research on IT business value (Kohli & Devaraj, 2003). More recently the focus of the IT business value research has been shifted to the impact of a firm's IT resources and capabilities on various performance outcomes. For example, a number of IS researchers have investigated the role of IT in a firm's innovation process and innovation performance (Ravichandran et al., 2017; Saldanha et al., 2017; Tarafdar & Tanriverdi, 2018; Trantopoulos et al., 2017). Moreover, the role of IT has also been examined in the context of mergers and acquisitions by IS researchers (Benitez et al., 2018). In addition, another stream of IT business value research has focused on the role of interorganizational IT resources and capabilities in value co-creation (Mandrella et al., 2020; Neirotti & Raguseo, 2017). In particular it has been found that IT intensity facilitates the knowledge flow between alliance partners (Ravichandran & Giura, 2019).

Given the increasingly strategic role of IT in contemporary organizations, IT spending has become a critical capital investment made by a firm (Turedi & Zhu, 2019). A large-scale IT investment often requires a more centralized arrangement of IT activities. Large scale investment projects are often of high-level investments involving mission critical systems such as enterprise applications or infrastructure deployment. For deploying such systems, a firm often needs a centralized coordination capability. For example, a centralized IS unit has been shown in the past to facilitate IT infrastructure development by providing the necessary coordination capabilities across the firm (Sambamurthy & Zmud, 2000). From a management perspective, managers in a centralized corporate-level IS unit are better positioned to identify opportunities for technology investments as well as the technical requirements for effective deployment of such investments (Xue et al., 2008). In the meantime, they are also equipped with the needed structural power and resources to overcome any impediment in the process of deploying a large-scale investment, particularly the potential resistance from local business units which may be more focused on their own objectives. With a more centralized IT capability for coordination, a large-scale IT investment is more likely to create value for the firm. Therefore, a more centralized arrangement of IT activities complements a large-scale investment in IT, which should lead to performance improvement.

Hypothesis 1: The centralization (decentralization) of IT activities is positively (negatively) associated with firm performance if the firm has a large-scale IT investment.

Firm Diversification

Firm diversification is an important corporate-level strategy. Previous research has shown that the scope of the firm has a significant impact on a number of IT activities such as IT investment (Liu & Ravichandran, 2008) and IT governance. In the context of IT governance, it has been shown that diversification mode and diversification breadth influence the choice of IT governance modes (Sambamurthy & Zmud, 1999). A diversified firm is often involved in two or more distinct businesses. As a result, the diverse nature of the different businesses presents a great managerial challenge. Strategic decision-making often yields better outcomes if the decision-making authority is delegated to the individual business unit level because managers at the business level often have first-hand information and knowledge about how to compete in a particular industry environment. In a diversified firm, accordingly, divisional managers or business-level managers are often assigned with the operating responsibility of conducting day-to-day activities in their respective product markets.

When considering a large and diversified firm's use of IT resources and capabilities, the focus on divisional affairs or business-level strategies would require a firm's IT resources and capabilities to concentrate on dealing with local technological needs, which requires a more decentralized organizational arrangement. In their study on business-IT alignment in multi-business organizations, Reynolds and Yetton (2015) argued that the close alignment between the firm's divisional strategy and divisional IT strategy would create value to the firm. More specifically the authors argue in this study that the alignment between a firm's diversification strategy and a more decentralized IS organization should yield a positive performance effect since the diversification strategy would benefit from an array of more localized IT capabilities to cope with various divisional needs. Therefore:

Hypothesis 2: The decentralization of IT activities is positively associated with firm performance if the firm has a high level of diversification.

Firms may pursue different types of diversification strategies. They may enter related industries that often share certain commonalities in value chain functions between them or they may enter unrelated industries that have no linkages to exploit between them. Related firms often seek for economies of scope through sharing activities or resources between business units, known as synergistic economies, whereas unrelated firms may take advantage of the internal capital market to allocate funds more efficiently than the external capital market, known as financial economies (Hill & Hoskisson, 1987). Relatedness in a firm's business portfolio often has an impact on the choice of IT governance modes. Firms that compete in related industries tend to have commonalities in their IT infrastructures and applications (Tanriverdi, 2005), which may require a centralized locus for IT-related decision making (Sambamurthy & Zmud, 1999). In other words, a centralized arrangement of IT activities in a related firm seems to provide the needed resources and capabilities for realizing economies of scope or synergistic economies. Hence a more centralized IS organization should complement a firm's related diversification strategy.

On the other hand, when firms compete in unrelated industries, the marked difference between the industries makes it difficult for realizing economies of scope. It also becomes more difficult for the staff in a centralized corporate-level IS unit to understand divisional business strategies and value creation processes due to the distance to the local conditions. Under this circumstance, the local and business-level IS staff members are better positioned to develop technologies and applications that meet the distinctive IT needs of business units (Sambamurthy & Zmud, 1999; Reynolds & Yetton, 2015). As a result, the IS organization should be structured in a way that accommodates the localization need for technology development through assigning business-level IT units autonomy and flexibility. Normally a more decentralized organizational arrangement of IT activities would fulfill such a need. Since each unrelated business unit essentially operates as an independent and autonomous entity, a decentralized IS organization should help each business unit develop complementary IT sources and capabilities and effectively compete in its respective industry, which should contribute to the overall performance of the firm. To sum up the above arguments, the authors propose the following hypotheses:

Hypothesis 3a: The centralization (decentralization) of IT activities is positively (negatively) associated with firm performance if the firm has a high level of related diversification.

Hypothesis 3b: The decentralization of IT activities is positively associated with firm performance if the firm has a high level of unrelated diversification.

METHODOLOGY

Data

The authors collected their data from the Compustat database and the annual surveys conducted by InformationWeek in their special issue InformationWeek 500. InformationWeek conducted annual surveys since 1989 on the innovative firms regarding their IT usage, which they published in the annual special issue named InformationWeek 500. The surveys contained items related to firms' usage of IT and provided valuable information for research purposes. The authors computed the variables for performance measures, diversification levels, and control variables from Compustat. IT investment and IS organization data were obtained from InformationWeek 500 during the 1999-2003 time period. The authors matched the firms listed in InformationWeek with firms in the Compustat database and their sample consisted of 798 firm-year observations during the five-year period.

Dependent Variable

The dependent variable in this study is firm profitability, measured by the return on sales (ROS). The ROS was computed by dividing a firm's net income by its total annual sales. The authors used data from Compustat for computing the ROS of the firms in the sample.

Independent Variables

The main independent variable, the organizational arrangement of IT activities (IT decentralization), was measured by the ratio of the firm's business IT employees to corporate IT employees. The data used to compute this variable was obtained from the InformationWeek annual surveys. A large value of this variable indicates a more decentralized IT organization whereas a small value indicates a more centralized approach to IT organization. The scale of IT investment was measured by the ratio of IT budget to the firm's sales revenue, which was also obtained from the InformationWeek surveys.

In this study the authors used the entropy measure to measure firm diversification, not only because it is widely accepted but also because as one of its key properties, it allows for the authors to distinguish between related and unrelated diversification (Palepu, 1985). Given the study's focus on related and unrelated diversification this measure is more appropriate than other available measures of product diversification. Following procedures used in past strategy and IS studies (Palepu, 1985; Ravichandran et al., 2009) the authors calculated the entropy measures for related and unrelated diversification using business segment data from Compustat. Total firm diversification was measured as the sum of related and unrelated diversification.

Control Variables

In the empirical analysis, the authors included dummy variables controlling for the fixed effect of industry sectors. Year dummies were also included since the observations ranged from 1999 to

2003. Since within the same industry environment firm performance varies due to the different strategies they pursue, firm characteristics such as size, R&D expenditure, advertising intensity, capital intensity, and financial leverage were included as controls in the analysis. The authors controlled for firm size measured by the natural logarithm of a firm's total assets (in billion dollars). Following past studies, the authors used R&D intensity, the ratio of R&D expenditures to the firm's total sales, as a control variable. The authors measured advertising intensity by dividing the advertising expenditure by total sales. Firm leverage was measured by the widely used ratio of long-term debt to total capital (debt plus equity). Capital intensity is measured as the percentage of capital expenditures to sales. Table 1 shows the summary statistics and correlations matrix.

Table 1. Summary Statistics and Correlation Matrix

	Variable	Mean	SD	1	2	3	4	5	6	7	8	9	10	11
1	ROS	0.156	0.148	1										
2	IT Decentralization	0.242	1.474	-0.02	1									
3	IT Investment	3.400	5.811	0.08*	-0.00	1								
4	Firm Diversification	0.604	0.462	0.06	0.05	0.04	1							
5	Related Diversification	0.209	0.302	0.03	-0.03	0.08*	0.57**	1						
6	Unrelated Diversification	0.395	0.380	0.05	0.09*	-0.02	0.76**	-0.10	1					
7	Firm Size	8.285	1.331	0.27**	-0.05	0.00	0.25**	0.20**	0.15**	1				
8	R&D Intensity	0.043	0.137	-0.63**	-0.02	0.05	-0.15**	-0.07†	-0.13**	-0.19**	1			
9	Advertising Intensity	0.011	0.031	0.09*	-0.02	0.04	-0.03	-0.02	-0.01	0.07*	-0.02	1		
10	Leverage	0.231	0.155	0.11**	-0.02	0.02	0.03	0.03	0.01	-0.08*	-0.20**	-0.04	1	
11	Capital Intensity	0.071	0.086	0.26**	-0.04	0.05	-0.00	0.04	-0.03	0.12**	0.02	-0.07†	0.18**	1

Note: † significant at 10%; * significant at 5%; ** significant at 1%.

ANALYSIS AND RESULTS

The dataset turned out to be a very imbalanced panel data with 798 observations over five years and 338 firms. Since on average only 2.36 observations were available for each firm during the five-year period, instead of employing an advanced time-series technique, the authors decided to use Ordinary Least Squares (OLS) regression for the pooled data. Also, the authors ran the analysis with Huber-White standard errors clustering on firms, which will produce a robust estimation of standard errors despite the presence of arbitrary correlations in error terms within a cluster (firm).

First the authors ran a model which included only control variables. The result was reported in Table 2 under Model 1. Model 2 includes the variable of IT investment and the variable of IT decentralization. Model 3 includes the interaction term between the IT investment variable and IT organization variable, which was used to test H1. By H1 the authors expected a negative (positive) interaction effect between IT investment and the decentralization (centralization) of IT employees. The coefficient of the interaction

Table 2. Regression Analysis of the Impact of IT Organizational Forms on Firm Performance							
	I	II	III	IV	V	VI	VII
R&D Intensity	-0.624** (0.154)	-0.661** (0.134)	-0.663** (0.133)	-0.684** (0.117)	-0.689** (0.115)	-0.684** (0.117)	-0.688** (0.116)
Advertising Intensity	0.299† (0.179)	0.262 (0.185)	0.258 (0.186)	0.316† (0.189)	0.306 (0.189)	0.314† (0.186)	0.305 (0.188)
Leverage	0.020 (0.045)	0.020 (0.045)	0.023 (0.045)	-0.049 (0.035)	-0.050 (0.035)	-0.048 (0.035)	-0.049 (0.035)
Capital Intensity	0.416† (0.251)	0.435† (0.258)	0.432† (0.258)	0.482** (0.154)	0.481** (0.153)	0.483** (0.154)	0.485** (0.151)
Size	0.011* (0.005)	0.012* (0.005)	0.013* (0.005)	0.015** (0.005)	0.015** (0.005)	0.015** (0.005)	0.015** (0.005)
IT Investment		0.004* (0.002)	0.005** (0.002)	0.003** (0.001)	0.003** (0.001)	0.003** (0.001)	0.003** (0.001)
IT Decentralization		-0.003 (0.003)	-0.008* (0.003)	-0.002 (0.002)	-0.013** (0.005)	-0.002 (0.002)	-0.017** (0.006)
IT Investment × IT Decentralization			-0.003* (0.001)				
Firm Diversification				-0.020 (0.013)	-0.019 (0.012)		
Diversification × IT Decentralization					0.014** (0.005)		
Related Diversification						-0.026† (0.016)	-0.031* (0.015)
Unrelated Diversification						-0.016 (0.015)	-0.014 (0.015)
Related Diversification × IT Decentralization							-0.019 (0.016)
Unrelated Diversification × IT Decentralization							0.014** (0.005)
Constant	0.230** (0.042)	-0.008 (0.049)	-0.013 (0.049)	-0.004 (0.047)	0.001 (0.047)	-0.006 (0.048)	-0.000 (0.048)
Observations	798	798	798	740	740	740	740
R-squared	47.6%	50.7%	51.1%	53.4%	53.7%	53.4%	53.8%
ΔR-squared			0.4%**		0.3%**		0.4%†
Note: 1. Robust standard errors are reported in parentheses; 2. Coefficients of sector dummies and year dummies are not shown in the table; 3. Firm diversification is the sum of related diversification and unrelated diversification; 4. The number of observations changed in Model IV because there were missing values for the diversification variables and the sample size was reduced to 740; 5. † significant at 10%; * significant at 5%; ** significant at 1%.							

term is negative and significant (b=-0.003, $p<0.05$), which suggests that a more decentralized (centralized) IT organization has a negative (positive) effect on firm profitability when the scale of IT investment is high. The authors plot the interaction effect in Figure 1. As the figure shows, when the level of IT investment is high, the degree of IT centralization has a positive effect on profitability, indicating that a more centralized IT organization has a positive effect on firm performance. When the level of IT investment is low, the degree of IT centralization has a slightly negative effect on return on sales. This result supports H1.

Figure 1: The moderating effect of IT investment scale

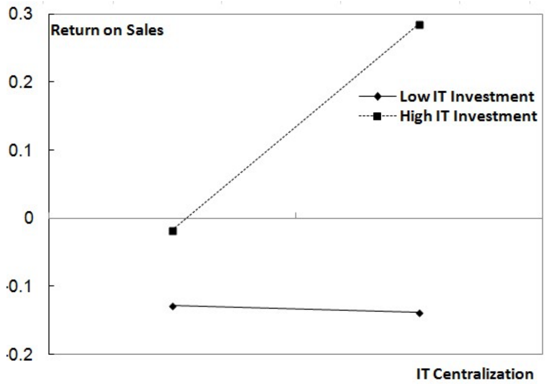
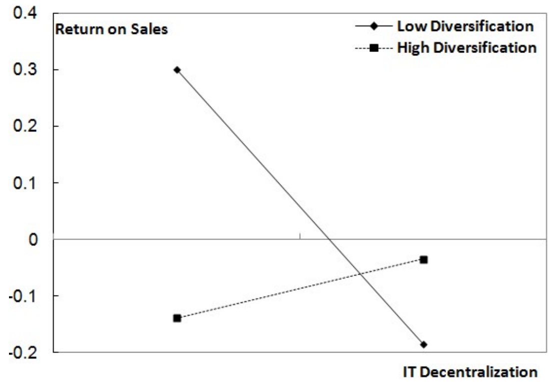
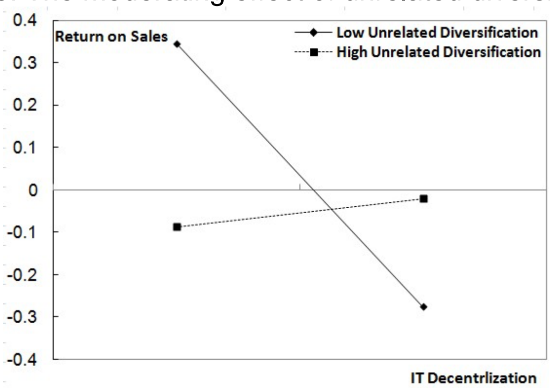


Figure 2: The moderating effect of firm diversification



Model 4 adds the variable of firm diversification (the sum of related and unrelated diversification), and Model 5 adds its interaction with IT decentralization to test H2. By H2, the authors expected to see a positive interaction effect and the result supports it ($b=0.014$, $p<0.01$), suggesting that when the level of diversification is high, IT decentralization has a positive effect on firm profitability. Likewise, the authors plot the interaction effect in Figure 2. It can be seen in the figure that the negative relationship between IT decentralization and return on sales turns positive as the level of diversification increases. H2 is therefore supported. To test H3a and H3b, the authors included the two interaction terms in Model 7: the interaction between IT decentralization and related diversification and the interaction between IT decentralization and unrelated diversification. By H3a the authors expected a negative interaction effect of related diversification and by H3b a positive interaction effect of unrelated diversification. The former interaction, as revealed by the result in Model 7, has a negative coefficient that is not statistically significant ($b=-0.019$, $p=0.248$). The latter has a positive and significant coefficient ($b=0.014$, $p<0.01$), supporting H3b. Figure 3 shows the interaction effect of unrelated diversification.

Figure 3: The moderating effect of unrelated diversification



DISCUSSION

Research on IT governance has long heeded the importance of organizing IT activities for IT capabilities (Sambamurthy & Zmud, 2000). As firms attempt to consistently reap above-average returns from their IT investments, an important condition is that they make IT-related decisions effectively on a consistent basis. To do so, they often need to have an appropriate yet flexible IT organizational form to administer all major IT decisions. These decisions can be related to the management of technology expenditures and investments, budgeting and approval of IT spending, and application development. Such important IT decisions entail the usefulness and effectiveness of the overall IT strategy of the firm, and therefore have a critical influence on the overall success of IT use in the firm. Appropriate IT organizational structures should facilitate the development of relevant and complementary IT resources and capabilities that lead to the maximization of firm efficiency and effectiveness. As a result, the choice regarding the IT organizational form should be a concerted effort to create firm-level capabilities enabled by IT.

This study adopts the aforementioned resource-based perspective and investigates the potential performance impacts of IT organizational forms under certain circumstances. As the authors found in this study, when the scale of IT investment is high, a more centralized form of IT activities positively affects firm performance, suggesting a strong coordination need in such firms. In addition, when the diversification level is high, the authors found that a decentralized form of IT activities often yields better performance outcomes, suggesting that in diversified firms there is a strong need to localize IT resources and capabilities to create custom applications and services. Similarly, a strategy of unrelated diversification also interacts with a decentralized IT organizational form to positively affect firm performance. However, the authors did not find a positive interaction effect between related diversification and IT centralization. The authors suspect that in a diversified firm, the localization need of unrelated diversification might suppress the coordination need of related diversification. Hence a more decentralized organizational arrangement of IT activities is the dominant form in such a firm and the effect of related diversification is therefore weakened. Although this paper did not find the positive interaction effect, there has been research showing that certain forms of IT governance might in fact contribute to both IT and business process relatedness (Kude et al., 2017). The authors hope that future research using different data sets or methodologies such as case studies can verify the suspected suppression effect and investigate the role of IT governance in multi-business firms further.

Previous research has shown that organization structures have a direct influence on IT governance forms (Ahituv et al., 1989). Diversification can affect the structure used by a multi-business firm to manage the complexity created by a diversified business portfolio (Hill, 1985). It has been argued that diversified firms often adopt a multidivisional structure to separate the strategic responsibilities at the corporate level and the operating responsibilities at the business level, although the structures used by diversified firms can differ to address the specific challenge imposed by different types or levels of diversification strategy (Hill et al., 1992). By the same token, IT governance is likely to be decentralized to address the demand of a multidivisional structure as firm diversification increases. The structure or form of IT organizations may thus complement the diversification strategy and enhances a firm's capability in using IT effectively, which eventually would enhance the firm's returns from IT investments.

This study also contributes to the broad literature on strategic IT alignment. After more than two decades of research on business-IT alignment, the subject has remained as an important phenomenon both in academic research and practitioner's society (Coltman et al., 2015). It has been argued and shown that the alignment between a firm's IT resources and capabilities and its business strategy would have a positive effect on firm performance (Tallon & Pinsonneault, 2011; Oh & Pinsonneault, 2007; Yayla & Hu, 2012). In particular, a recent study (Wu et al., 2015) has shown that IT governance mechanisms measured by decision-making structure, formal processes, and communication approach had a positive impact on business-IT alignment and subsequently organizational performance. In addition, prior research has shown that the alignment between a firm's corporate strategy and complementary IT resources and capabilities would also affect firm performance (Broadbent et al., 1996; Hodgkinson, 1996; Reynolds & Yetton, 2015). The findings of the study could also be viewed through this conceptual lens in the sense that the different IT organizational arrangements appeared to be complementary to a firm's diversification strategies, perhaps by enhancing the firm's strategic IT alignment.

The results from this study may also suggest that the business value derived from IT investment might depend on how a firm's IT organization is structured. In the context of a large-scale IT investment, the firm is more capable of creating value from IT by having a more centralized IT governance form since centralized decision-making often leads to greater efficiency, easier coordination, and consistency in the firm's IT strategy. As a result, this study also contributes to the much broader IT business value research. Despite decades of research in this area, unfortunately it seems that managers are still questioning the business value of IT (Turedi & Zhu, 2019). This study, along with other IT business value studies, has shown the importance of considering the different dimensions of the IT variable (Benitez et al., 2018; Kim et al., 2018;) as well as the synergistic effect between IT resources and capabilities and complementary organizational characteristics (Kohli & Devaraj, 2003; Nwankpa & Datta, 2017).

CONCLUSION

This study contributes to the broad IT governance research by establishing the linkage between a firm's IT organizational form and firm performance. The authors have shown that the linkage is contingent upon certain characteristics such as the firm's IT investment scale and diversification strategy. The finding may suggest that there exist potential synergistic effects on firm performance of IT governance and a firm's strategic choices. The authors hope that their findings will encourage future researchers to explore the matter in more depth by identifying other contingencies, particularly those stemming from the firm's external environment. For IT practitioners, the findings of this study would offer them valuable insights regarding how to structure IT organizations and prescriptions and find out which type of IT organizational forms fit their particular internal environment the best. The existing literature on IT governance lacks an in-depth consideration of strategic variables such as diversification and this paper attempts to fill this gap. The findings suggest that policy makers in the firm should simultaneously focus on various factors such as IT governance, IT investment, and firm diversification as a whole when seeking economic value from IT. A synergistic view of IT governance forms and firm-level factors may yield better outcomes in terms of firm performance.

REFERENCES

Ahituv, N., Neumann, S., & Zviran, M. (1989). Factors Affecting Policy for Distributing Computing Resources. MIS Quarterly, 13(4), 388-402.

Ali, S., & Green, P. (2012). Effective information technology (IT) governance mechanisms: An IT oursourcing perspective. Information Systems Frontiers, 14(2), 179-193.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. Journal of Management, 17(1), 99-120.

Benitez, J., Llorens, J., & Braojos, J. (2018). How information technology influences opportunity exploration and exploitation firm's capabilities. Information & Management, 55(4), 508-523.

Benitez-Amado, J., Ray, G., & Henseler, J. (2018). Impact of information technology infrastructure flexibility on mergers and acquisitions. MIS Quarterly, 42(1), 25-43.

Broadbent, M., Weill, P., Brien, T., & Neo, B.-S. (1996). Firm Context and Patterns of IT Infrastructure Capability. ICIS 1996 Proceedings (pp. 174-194). Cleveland, OH: International Conference on Information Systems.

Brown, A. E., & Grant, G. G. (2005). Framing the Frameworks: A Review of IT Governance Research. Communications of the Association of Information Systems, 15, 696-712.

Chi, M., Zhao, J., George, J. F., Li, Y., & Zhai, S. (2017). The influence of inter-firm IT governance strategies on relational performance: The moderation effect of information technology ambidexterity. International Journal of Information Management, 37(2), 43-53.

Chong, J. L., & Duong, L. N. (2017). Understanding IT Governance Effectiveness in Asia: An Event Study. Pacific Asia Journal of the Association for Information Systems, 9(1), 29-54.

Coltman, T., Tallon, P., Sharma, R., & Queiroz, M. (2015). Strategic IT alignment: twenty-five years on. Journal of Information Technology, 30, 91-100.

Dewan, S., Michael, S. C., & Min, C.-k. (1998). Firm Characteristics and Investments in Information Technology: Scale and Scope Effects. 9(3), 219-232.

Grant, R. M. (1991). The resource-based theory of competitiveadvantage: implications for strategyformulation. California Management Review, 33(3), 114-135.

Gregory, R. W., Kaganer, E., Henfridsson, O., & Ruch, T. J. (2018). IT Consumerization and the Transformation of IT Governance. MIS Quarterly, 42(4), 1225-1253.

Henderson, J., & Venkatraman. (1993). Strategic Alignment: Leveraging information technology for transforming organizations. IBM Systems Journal, 32(1), 4-16.

Héroux, S., & Fortin, A. (2018). The moderating role of IT-business alignment in the relationship between IT governance, IT competence, and innovation. Information Systems Management, 35(2), 98-123.

Hill, C. W. (1985). Oliver Williamson and the M-form firm: a critical review. Journal of Economic Issues, 19(3), 731-751.

Hill, C. W., & Hoskisson, R. E. (1987). Strategy and Structure in the Multiproduct Firm. Academy of Management Review, 12(2), 331-341.

Hill, C. W., Hitt, M., & Hoskisson, R. (1992). Cooperative versus Competitive Structures in Related and Unrelated Diversified Firms. Organization Science, 3(4), 501-521.

Hodgkinson, S. L. (1996). The Role of the Corporate IT Function in the Federal IT Organisation. In M. J. Earl, & M. J. Earl (Ed.), Information Management: The Organizational Dimension. Oxford: Oxford University Press.

Huber, T. L., Kude, T., & Dibbern, J. (2017). Governance Practices in Platform Ecosystems: Navigating Tensions Between Cocreated Value and Governance Costs. Information Systems Research, 28(3), 563-584.

Kim, T. H., Wimble, M., & Sambamurthy, V. (2018). Disaggregation of the IT capital effects on firm performance: Empirical evidence from an IT asset portfolio perspective. European Journal of Information Systems, 27(4), 449-469.

Kohli, R., & Devaraj, S. (2003). Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research. Information Systems Research, 14(2), 127-145.

Kude, T., Lazic, M., Heinzl, A., & Neff, A. (2017). Achieving IT-based synergies through regulation-oriented and consensus-oriented IT governance capabilities. Information Systems Journal, 28(5), 765-795.

Liang, H., Wang, N., Xue, Y., & Ge, S. (2017). Unraveling the Alignment Paradox: How Does Business-IT Alignment Shape Organizational Agility? Information Systems Research, 28(4), 863-879.

Liu, Y., & Ravichandran, T. (2008). A Comprehensive Investigation on the Relationship between Information Technology Investments and Firm Diversification. Information Technology and Management, 9(3), 169-180.

Mandrella, M., Trang, S., & Kolbe, L. M. (2020). Synthesizing and Integrating Research on IT-Based Value Cocreation: A Meta-Analysis. Journal of the Association for Information Systems, 21(2), 388-427.

Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. MIS Quarterly, 28(2), 283-322.

Neirotti, P., & Raguseo, E. (2017). On the contingent value of IT-based capabilities for the competitive advantage of SMEs: Mechanisms and empirical evidence. Information & Management, 54(2), 139-153.

Nwankpa, J., & Datta, P. (2017). Balancing exploration and exploitation of IT resources: the influence of Digital Business Intensity on perceived organizational performance. European Journal of Information Systems, 26(5), 469-488.

Oh, W., & Pinsonneault, A. (2007). On the Assessment of the Strategic Value of Information Technologies: Conceptual and Analytical Approaches. MIS Quarterly, 31(2), 239-265 .

Palepu, K. (1985). Diversification Strategy, Profit Performance and the Entropy Measure. Strategic Management Journal, 6(3), 239-255.

Park, J., Lee, J.-N., Lee, O.-K. D., & Koo, Y. (2017). Alignment Between Internal and External IT Governance and Its Effects on Distinctive Firm Performance: An Extended Resource-Based View. IEEE Transactions on Engineering Management, 64(3), 351-364.

Queiroz, M. (2017). Mixed results in strategic IT alignment research: a synthesis and empirical study. European Journal of Information Systems, 26(1), 21-36.

Ravichandran, T., & Giura, S. I. (2019). Knowledge Transfers in Alliances: Exploring the Facilitating Role of Information. Information Systems Research, 30(3), 726-744.

Ravichandran, T., & Liu, Y. (2011). Environmental factors, managerial processes, and information technology investment strategies. Decision Sciences, 42(3), 537-574.

Ravichandran, T., Han, S., & Mithas, S. (2017). Mitigating Diminishing Returns to R&D: The Role of Information Technology in Innovation. Information Systems Research, 28(4), 812-827.

Ravichandran, T., Liu, Y., Han, S., & Hasan, I. (2009). Diversification and Firm Performance: Exploring the Moderating Effects of Information Technology Spending . Journal of Management Information Systems, 25(4), 205-240.

Raymond, L., Bergeron, F., Croteau, A.-M., & Uwizeyemungu, S. (2019). Determinants and outcomes of IT governance in manufacturing SMEs: A strategic IT management perspective. International Journal of Accounting Information Systems, 35.

Reynolds, P., & Yetton, P. (2015). Aligning business and IT strategies in multi-business organizations. Journal of Information Technology, 30(2), 101–118.

Sabherwal, R., Sabherwal, S., Havakhor, T., & Steelman, Z. (2019). How does strategic alignment affect firm performance? the roles of information technology investment and environmental uncertainty. 43(2), 453-474.

Saldanha, T. J., Mithas, S., & Krishnan, M. S. (2017). Leveraging Customer Involvement for Fueling Innovation: The Role of Relational and Analytical Information Processing Capabilities. MIS Quarterly, 41(1), 367-396.

Sambamurthy, V., & Zmud, R. W. (1999). Arrangements for Information Technology Governance: A Theory of Multiple Contingencies. MIS Quarterly, 23(2), 261-290.

Sambamurthy, V., & Zmud, R. W. (2000). Research Commentary: The Organizing Logic for an Enterprise's IT Activities in the Digital Era - A Prognosis of Practice and a Call for Research. Information Systems Research, 11(2), 105-114.

Tallon, P. P., & Pinsonneault, A. (2011). Competing Perspectives on the Link Between Strategic Information Technology Alignment and Organizational Agility: Insights from a Mediation Model. MIS Quarterly, 35(2), 463-486.

Tanriverdi, H. (2005). Information Technology Relatedness, Knowledge Management Capability, and Performance of Multibusiness Firms. MIS Quarterly, 29(2), 311-334.

Tarafdar, M., & Tanriverdi, H. (2018). Impact of the Information Technology Unit on Information Technology-Embedded Product Innovation. Journal of the Association for Information Systems, 19(8), 716-751.

Trantopoulos, K., von Krogh, G., Wallin, M. W., & Woerter, M. (2017). External knowledge and information technology: Implications for process innovation performance. MIS Quarterly, 41(1), 287-300 .

Turedi, S., & Zhu, H. (2019). How to Generate More Value from IT: The Interplay of IT Investment, Decision Making Structure, and Senior Management Involvement in IT Governance. Communications of the Association for Information Systems, 44, 511-536.

Turel, O., Liu, P., & Bart, C. (2017). Board-Level Information Technology Governance Effects on Organizational Performance: The Roles of Strategic Alignment and Authoritarian Governance. Information Systems Management, 34(2), 117-136.

Weill, P. (2004). Don't Just Lead, Govern: How Top-Performing Firms Govern IT. MIS Quarterly Executive, 3(1), 1-17.

Wernerfelt, B. (1984). A Resource-Based View of the Firm. Strategic Management Journal, 5(2), 171-180.

Wu, S. P.-J., Straub, D. W., & Liang, T.-P. (2015). How information technology governance mechanisms and strategic alignment influence organizational performance: Insights from a matched survey of business and IT managers. MIS Quarterly, 39(2), 497-518.

Xue, L., Ray, G., & Gu, B. (2011). Environmental Uncertainty and IT Infrastructure Governance: A Curvilinear Relationship. Information Systems Research, 22(2), 389-399 .

Xue, L., Ray, G., & Zhao, X. (2017). Managerial Incentives and IT Strategic Posture. Information Systems Research, 28(1), 180-198.

Xue, Y., Liang, H., & Boulton, W. R. (2008). Information Technology Governance in Information Technology Investment Decision Processes: The Impact of Investment Characteristics, External Environment, and Internal Context. MIS Quarterly, 32(1), 67-96.

Yayla, A. A., & 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, 21(4), 373-387.

DECISION SCIENCES INSTITUTE

Examining the drivers of Supply Chain Integration: a test of competing theories and models

Benjamin Agyei-Owusu
Kwame Nkrumah University of Science and Technology
Email: bagyei-owusu.ksb@knust.edu.gh

Ishmael Nanaba Acquah
Kwame Nkrumah University of Science and Technology / University of Education, Winneba
Email: inacquah1@st.knust.edu.gh / inacquah@uew.edu.gh

David Asamoah
Kwame Nkrumah University of Science and Technology
Email: dasamoah.ksb@knust.edu.gh

Bernice Edem Asemsro
Kwame Nkrumah University of Science and Technology
Email: beasemsro@st.knust.edu.gh

ABSTRACT

The research examines the drivers of supply chain integration (SCI) from the competing theoretical perspectives of the relational view and resource-based view theories. Two competing models - a relationship factors model (model 1) and an organizational factors model (model 2) are developed and examined using data from 83 manufacturing and service firms. The findings revealed that two relationship drivers (trust and commitment) and two organizational drivers (top management support and goal compatibility) had positive and significant effects on SCI. Overall, the organizational factors model was found to be stronger in predicting the level SCI. Implications of the findings were discussed.

KEYWORDS: Supply Chain Integration, competing theories, partial least squares structural equation modelling, Survey research

INTRODUCTION

In the present-day business environment, working closely with supply chain partners is vital to firm survival (Hendijani & Saei, 2020). Supply chain integration (SCI) is the extent to which a firm interacts and collaborate with its other supply chain partners to achieve efficient and effective flows of information, products, decisions, money and information in the most effective and efficient way (Zhao et al., 2008). Integration with supply chain partners enhances the level of an organization's services and performance. The objective of SCI is to provide high level of satisfaction and value to customers through speed and cost-effective flows of information and materials (Flynn et al., 2010). SCI is generally acknowledged to be vital in attaining high business performance and competitive advantage, thus it is unsurprising that supply chain management literature is replete with SCI studies (Huo et al., 2014).

The focus of the extensive exploration of SCI in literature has however centred on performance outcomes such as operational (Lu et al., 2017; Flynn et al., 2010; Yuen & Thai, 2016), financial (Zhao et al., 2015), firm performance (Liu & Wei, 2011; Huo et al., 2014; Hendijani & Saei, 2020). Research on the drivers of SCI is comparatively less developed, and there have been calls for more research on drivers of SCI (Xu et al., 2014; Flynn et al., 2021). It is currently not clear from the extant literature what factors drive firms to practice SCI and the factors that account for different levels of SCI among organizations. This paper seeks to address this research gap by conducting an intensive examination of the drivers of SCI using the competing theories and models approach. A review of supply chain and organizational behaviour literature by Lee et al. (2010) found that key relationship drivers (trust, commitment and interdependency) and organizational drivers (top management support, cultural similarity and goal compatibility) may help explain supply chain activities of firms. According to Lee et al. (2010), the recognition of the significance of these antecedents is necessary for successful supply chain management. The extent to which firms practice SCI, an important supply chain activity, could thus be explained by relationship and organizational drivers. The present study adopts the competing models approach to examine the drivers of SCI from these two competing perspectives. The competing models approach is unique and especially useful where there are multiple alternative explanations for a phenomenon (Mital et al., 2018).

The rest of the paper is structured as follows. The next section focuses on the literature review, theoretical background, research model and hypotheses development. The focus of the subsequent section is the study’s methodology followed by presentation and discussion of results. The study concludes by discussing the implications of the study in addition to providing directions for further research.

LITERATURE REVIEW

Supply Chain Integration

SCI can be characterized as the degree to which a company can collaborate strategically with its supply chain partners and manage intra- and inter-organizational processes in a cooperative manner to achieve effective and efficient flows of products, services, information, money, and decisions to provide the final customer with maximum value at low cost and high speed (Van Der Vaart and Van Donk, 2008). The goal of SCI is to build streamlined supply chains with completely synchronized upward knowledge and downstream flow of materials (Danese et al., 2013). There are two fundamental types of SCI namely internal integration and external integration. Linking internal processes to external vendors and consumers is a requirement for progress and there is a consensus among scholars on the strategic value of combining intra- and inter-organizational operations (Vickery et al., 2013). SCI is considered as one of the most vital activities to leverage both internal and external networks (He and Lai, 2012; Cousins and Menguc, 2006).

Internal integration is the organizational practice of combining and enhancing internal resources and information in order to generate knowledge sharing beyond individual functions or departmental boundaries to aid external integration efforts and to attain organizational goals (Zhao et al., 2011; Koufteros et al., 2010). Also, it entails different functional teams with the required expertise to share their skills, information and take good decisions about the organization’s operational processes together. (Koufteros et al., 2005). Increase in the level of internal integration can aid organizations to appropriately harmonize resources within the organization to improve capabilities across a lot of areas like quality, performance, cost, and delivery (Verona, 1999). *Supplier integration* is referred as the organizational practice where a buying firm and its suppliers share and apply operational, financial as well as strategic knowledge with the aim of generating mutual benefits (Narasimhan et al., 2010; Swink et al., 2007). Supplier integration aids in meeting customer requirements for products by providing unity of effort (Narasimhan & Kim, 2002) in addition to responding to changes in the marketplace (Zhao *et al*., 2013). *Customer integration* represents “close collaboration and information sharing activities with key customers that provide the firm with strategic insights into market expectations and opportunities, ultimately enabling a more efficient and effective response to customer needs” (Schoenherr & Swink, 2012; p.100). Customer integration provides a holistic understanding of market expectations and opportunities and enhances the ability to respond to customer needs and requirements in a more accurate and swift manner by ensuring a match between supply and demand (Swink et al., 2007).

Relationship drivers of supply chain integration

Trust

Capaldo and Giannoccaro (2015) assert that trust is important to supply chain management and trust has been found to be a significant predictor of positive performance outcomes in supply chain contexts. Trust in the context of supply chain acts an informal governance mechanism that enhances organizational coordination (Capaldo, 2014). Zaheer (1998) consider trust from the organizational level as the degree to which members of an organization have a collectively held trust orientation towards partner firms. Trust is beneficial in inter-firm exchanges as it creates an environment where firms endeavour to exceed the minimum requirements of a relationship to enhance the likelihood of mutual benefits (Panayides and Lun, 2009).

Commitment

Besides trust, commitment remains a vital relationship marketing construct which highlights the willingness of trading partners to take a long-term orientation of their relationship (Ganesan, 1994; Noordewier et al., 1990). Thus, commitment is referred as the degree to which supply chain partners will maintain and strengthen their business relationship (Lee et al., 2010). Commitment as informal self-enforcing governance is useful in generating relational rents and competitive advantages by minimizing the costs of transactions while maximizing value-creation initiatives (Yuan et al., 2018).

Interdependency

The primary concern of supply chain management is how independent players can be coordinated to work together to pursue the basic goal of profitability in changing market conditions (Simatupang and Sridharan, 2002). Interdependency refers to a firm’s need to maintain its business relationships with its supply chain actors to accomplish its goals and reflects the bargaining power and the importance of relationship among partners (Mahapatra et al., 2010). In supply chain, interdependency relate to the degree to which supply chain relationship participants’ processes is dependent on each other to accomplish their goals and realize the overall value creation (Simatupang et al., 2002; Capaldo and Giannoccaro, 2015)

Organizational drivers of supply chain integration

Top Management Support (TMS)

TMS relate to the willingness of senior managers to contribute time and resources to supply chain management activities that nurture and develop long-term relationships (Lockstrom et al., 2010). Thus, TMS can be considered as time and resources devoted by top management to the development of SCI (Zhao et al., 2015). TMS influences the ability to integrate information sharing strategy into the overall business strategy of an organization and the devoting of resources to improve information processing capabilities (Li and Lin, 2006).

Cultural Similarity

Cultural similarity entails the ability of an organization and its partners to have similar norms, values, beliefs in the supply chain. It helps maintain consistency of business practices (McAfee et al., 2002). Cultural similarity is a vital feature that distinguishes long-term business-to-business partnerships from other short-term alliances (Rajaguru & Jekanyika, 2012). Having similar or compatible culture enhances effective communication as well as knowledge exchange leading to fostering of shared goals (Emden et al., 2006).

Goal Compatibility

Goal compatibility is referred as the degree to which supply chain partners have clear and agreed-upon transactional goals (Lee et al., 2010). It is important for an organization to integrate with partners in the chain that share the same or similar business goals. Without a compatible goal, it will be difficult to achieve synergy.

THEORETICAL BACKGROUND

Two competing theories, namely the relational view and resource-based view (RBV), are adopted to examine the effect of the relationship antecedents (trust, commitment and interdependency) and organizational antecedents (top management support, cultural similarity and goal congruence) on SCI. The relational view serves as the theory explaining the hypothesized relationship antecedents of SCI while the RBV serves as the theory explaining the hypothesized organizational antecedents of SCI.

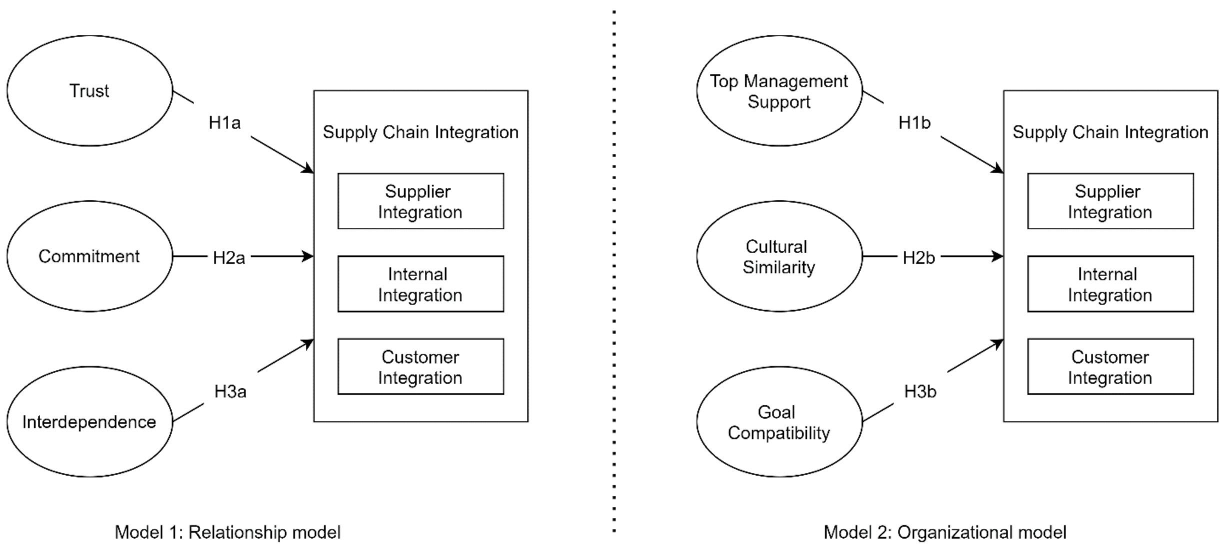
The underlying tenet of the relational view is that “a pair or network of firms can develop relationships that result in sustained competitive advantage” (Dyer and Singh, 1998: 675). It is challenging to obtain competitive advantages from within the boundaries of a firm only since important resources may span beyond the firm’s organizational boundary (Dyer and Singh, 1998). The relational view considers four sources that can generate inter-organizational competitive advantages (Dyer and Singh, 1998). These are (1) investments in relation-specific assets, (2) knowledge-sharing routines, (3) combination of complementary resources/capabilities, and (4) effective governance. Thus, the model grounded on the relational view (Model 1) examines how supply chain partners leverage on internal and external supply chain relationships for competitive advantage. Additionally, SCI normally extends beyond the boundaries of a single organization to include suppliers and customers resulting in an interdependence on other actors in the supply chain. Therefore, the relationship between integrated actors must be governed carefully for favourable performance outcomes. The relationship model proposes that trust, commitment and interdependency enable firms to obtain relational benefits from both internal and external supply chain partners for higher competitive advantage in terms of SCI.

In addition, the RBV considers organizations as collection of resources comprising of physical, human and organizational resources (Barney, 1991; Amit and Shoemaker, 1993) in addition to capability and competencies (Yang et al., 2009). The emphasis of the RBV is on the conceptualization of difficult-to-imitate attributes of the firm as sources for achieving superior performance and competitive advantage (Barney, 1986; Hamel and Prahalad, 1996). Such superior performance can only be achieved when a firm’s resources are valuable, rare, inimitable and non-substitutable, thereby constituting a strategic resource (Barney, 1991). The RBV of the firm suggests that unique and hard-to-imitate capabilities that contribute to superior performance are obtained from bundling the tangible and intangible resources that organizations control (Barney, 1991; Dangayach and Deshmukh, 2001). In line with the RBV, the study considers the organizational antecedents of top management support, cultural similarity and goal compatibility as a bundle of intangible resources which when leveraged can crucially influence SCI implementation and outcomes. Top management support are valuable critical resources needed by firms to manage their supply chains (Xu et al., 2014) including initiatives such as SCI. Additionally, the RBV considers organizational culture, herein cultural similarity, as a valuable intangible resource that impacts the effectiveness of organizations' capabilities to accomplish organizational objectives such as SCI (Eddleston et al. 2008; Bacq and Eddleston, 2018). The unique ways firms can bundle the organizational antecedents cannot be easily transferred to other firms without cost and can serve as a basis for sustained advantage.

RESEARCH MODEL AND HYPOTHESIS

Based on the two competing theories, two competing models of drivers of SCI were developed and tested in this study. Model 1, the relationship model, proposes that SCI is driven by three relationship-based antecedents namely trust, commitment and interdependence. Model 2, the organizational model, proposes that SCI is driven by three organizational antecedents namely top management support, cultural similarity and goal compatibility. The two models, presented in Figure 1 below, are analysed and compared to obtain insights into drivers of SCI. The hypothesized relationships for the models are discussed next.

Figure 1: Research models



Trust and SCI

Trust relates to the degree to which firms believe that supply chain partners will perform their obligations to each other in good faith (Lee et al., 2010). SCI, which involves internal and external integration activities, requires trust between internal functions and also external supply chain actors (Huo, 2012). The study considers trust as a relational factor that will facilitate SCI from the relational view perspective. Trusting organizations have the willingness of collaborating with their trusted partners with the expectation that those partners will not act opportunistically

(Lai et al., 2012). For example, Lee et al. (2010) indicate that firms will share information with the belief that other partners in the supply chain will forego opportunistic behaviour and utilize the information shared in a way that is mutually beneficial. Thus, trust can be considered as vital to SCI and information sharing. Simchi-levi and Simchi-levi (2007) consider trust as important to building cross-boundary integration. Similarly, Vijayasarathy (2010) found trust to have positive effect on supply integration. Based on the above, the study hypothesizes that;
H1a: Trust has a positive effect on SCI

Commitment and SCI

Dion et al. (1992) as cited in Vijayasarathy (2010) indicate that commitment can encourage investments in resources, in addition to integration of processes and information sharing for sustained benefit of all partners in a supply chain. Information sharing is fundamental to SCI initiatives, therefore, there is the need for firms to spend money, time and labour in developing information systems for information exchange and to minimize risk because of information exposure (Lee et al., 2010). Therefore commitment is important to SCI initiative as it opens up partners' behaviour to share timely and accurate information with the aim of maintaining the supply chain relationship (Tsanos and Zografos, 2016). From the relational view perspective, commitment is a vital relational factor for developing and maintaining relationships among supply chain partners and for SCI effort. Vijayasarathy (2010) found commitment to have a significant effect on supply integration while Zhao et al. (2008) found relationship commitment to be associated with customer integration. Based on the above, the study hypothesizes that;
H2a: Commitment has a positive effect on SCI

Interdependency and SCI

Interdependency induces cooperative goals between firms and their suppliers and distributors leading to trusting and continuously enhanced relationships (Wong et al., 2005). Interdependency can be considered vital to SCI implementation because SCI necessitates the extension of activities beyond the boundaries of a single firm. According to Seggie et al. (2006), partner dependence boosts inter-firm system integration. When firms are interdependent, e.g. for product supply, process accomplishments, or the usage of shared resources, interaction occurs among them because of the need to coordinate their activities (Capaldo and Giannoccaro, 2015). Such interdependency has the potential to influence SCI initiatives. Based on the above, the study hypothesizes that;
H3a: Interdependency has a positive effect on SCI

Top Management Support and SCI

Simchi-levi and Simchi-levi (2007) consider top management commitment as a centralized decision-making system which is required in the building of cross-boundary integration. Top management support is considered as a scarce and valuable resource (Boonstra, 2013) that plays an important role in supply chain management (Krause, 1999). Therefore, TMS from the RBV perspective will constitute a valuable resource that will influence firm's SCI initiatives and outcomes. TMS as a social link when configured synchronously with information technology influences the integrative capability of firm's supply chain (Xu et al., 2014). Basically, top-level supply chain managers with the awareness of the need to achieve competitive advantage in the marketplace have the responsibility of undertaking initiatives such as SCI. TMS has the potential of solving conflicts on goals and resources stemming from SCI (Zhao et al., 2015). TMS is also required to create a favourable environment and develop management styles which

are vital to accomplishing the benefits of SCI implementation (Zhao et al., 2015). Xu et al. (2014) found that TMS has a significant effect on supplier and customer integration. Based on the above, the study hypothesizes that;
H1b: Top management support has a positive effect on SCI

Cultural Similarity and SCI

Cultural similarity as a valuable resource from the RBV perspective will ensure easy fit and alignment of inter-firm activities related to SCI initiatives and subsequently contribute to the effectiveness of SCI. Rich (2003) indicate that partnering firms that have similar objectives and cultures can easily accomplish successful business-to-business partnerships and inter-organizational integration. Cultural similarity can therefore serve as an intangible and valuable asset that can influence the goal congruence of supply chain partners and influence interfirm activities such SCI, thereby, duly reinforced by the RBV. Cultural similarity (compatibility) is valuable because differences in the values and norms of partnering firms can negatively impact on inter-organizational information management and also information transfer and data-based decision making processes (Jones et al., 2005) and possibly affecting the effectiveness of SCI. Rajaguru and Jekanyika (2012) found that cultural compatibility has a positive impact on supply chain capabilities. Based on the above, the study hypothesizes that;
H2b: Cultural similarity has a positive effect on SCI

Goal Compatibility and SCI

Wathne and Heide (2000) stress that sharing common goals will inform cooperative activities between supply chain partners and enhance resource integration and utilization. Incongruence or differences in terms of supply chain partners' goals, objectives, information technology and culture possibly impede integration process (Rajaguru & Jekanyika, 2012). Having compatible goals with supply chain partners helps in understanding partner activities better leading to increased information and collaboration (Lee et al., 2010). Similarly, compatibility of goals and objectives of supply chain partners is critical in enhancing inter-organizational relationships and firm performance (Samaddar et al., 2006). Enhanced inter-organizational relationships stemming from goal compatibility can influence the implementation and effectiveness of SCI. Based on the above, the study hypothesizes that;
H3b: Goal compatibility has a positive effect on SCI

METHODOLOGY

The study examined the drivers of SCI by testing competing theories and models. Specifically, the study examined the effect of relationship drivers and organizational drivers on SCI using the competing theoretical perspectives of the relational view and the RBV theory. The items used for measuring the two drivers and SCI were adopted from prior literature. Items measuring relationship and organizational drivers were adopted from Lee et al. (2010) whereas the items measuring SCI were adopted from Wong et al. (2011). The items were all measured using a 5-point Likert scale anchored on 1 = strongly disagree and 5 = strongly agree. Additionally, demographic data such as firm size and industry type were collected. The measurement items were given to two supply chain academics and two industry players to ensure that the items fit the context of the study.

The target of the survey was manufacturing and service firms that operate in Ghana. A list of 1000 functioning manufacturing and service firms which have been in operation for more than a year was obtained from the Registrar General's Department to serve as the study's population. Out of this population, 250 firms were selected randomly for the collection of the study's data. Questionnaires served as the instrument of data collection and as a firm-level study, was issued to senior level managers of the sampled firms. After two rounds of follow-up, useable responses that were successfully retrieved were 83 representing a response rate of 33.2%. To test for the adequacy of the sample size, an a priori power analysis was conducted. Using a recommended α of 0.05 and a medium effect size of 0.15, and taking into account the three predictors of the dependent variable (SCI) for each separate model, the power test revealed that a minimum sample size of 76 will be enough to realize the recommended minimum power of 0.8. Therefore the 83 responses received was sufficient for valid results to be obtained. Non-response bias was tested by comparing the responses of the first fifteen respondents with the last fifteen respondents. The statistical test signified that there were no substantial differences between the initial fifteen respondents and the last fifteen respondents, indicating that non-response bias was not a problem.

RESULTS

The demographic analysis revealed that 16(19.3%) were state owned firms with the remaining 67(80.7%) being privately owned firms. In terms of the forms of ownership, 11(13.3%) were sole proprietorship and limited liability constituted 47(56.6%). Further, the industry type revealed majority of the firms were manufacturing firms (18.07%) and retail firms 13(15.66%). The annual revenue analysis revealed that 33(39.8%) firms had annual revenues between 5,000,001 and 20,000,000 million Ghana cedis with 17(20.5%) having annual revenues exceeding 20 million Ghana cedis. The analysis suggest that fairly large companies were sampled for the study as 81.9% of the firms had annual revenues exceeding one million Ghana cedis. The full demographic data is presented in Table 1.

Table 1: Demographic Data

Organization type	**Frequency**	**Percent**
State owned	16	19.3%
Privately owned	67	80.7%
Total	83	100
Organization Ownership	**Frequency**	**Percent**
Sole Proprietorship	11	13.3%
Limited Liability	47	56.6%
Partnership	7	8.4%
Public Limited Liability	17	20.5%
Total	83	100
Industry	**Frequency**	**Percent**
Manufacturing	15	18.1%
Financial Services (Banking & Investments)	9	10.8%
Oil and Gas	3	3.6%
Health	9	10.8%
Retail	13	15.7%
Construction	5	6.0%
Transportation	6	7.2%
Telecommunication	4	4.8%
Electronics and Computing Machinery	7	8.4%
Pharmaceuticals	7	8.4%

Automobile and Heavy Equipment	5	6.0%
Total	83	100
Approximate annual revenue in Ghana Cedis	**Frequency**	**Percent**
<40,000	3	3.6%
40,000-80,000	1	4.8%
80,001-200,000	5	6.0%
200,001-1,000,000	6	7.2%
1,000,001-5,000,000	18	21.7%
5,000,001-20,000,000	33	39.8%
>20,000,000	17	20.5%
Total	83	100

Measurement Model Results

The two competing research models developed in the study were analysed using the partial least squares structural equation modelling (PLS-SEM). The PLS-SEM results for the measurement model was first presented followed by the structural model results. To ensure that the model appropriately satisfied the recommended thresholds, relevant tests were performed. First, item loadings were examined to ensure that they loaded highly on their constructs. All the items were found to satisfy the minimum threshold of 0.708 as recommended (Hair et al., 2019). Also, the examination of the internal consistency of the construct revealed that all the items meet the composite reliability threshold of values above 0.7 as recommended (Hair et al., 2019). The evaluation of the AVE also revealed AVE values greater the recommended threshold of 0.5 thereby signifying acceptable convergent validity. Lastly, rho_A which is a consistent measure of PLS constructs satisfied the minimum threshold of 0.7 (Dijkstra and Henseler, 2015). As can be seen in Table 2, the item loadings, composite reliability, AVE and rho_A all satisfied the minimum thresholds.

Table 2: Psychometric properties of research constructs

Construct	Cronbach's Alpha	rho_A	Composite Reliability	AVE
Internal Integration	0.973	0.973	0.980	0.925
Customer Integration	0.972	0.972	0.978	0.899
Supplier Integration	0.967	0.967	0.974	0.883
Trust	0.964	0.965	0.974	0.902
Commitment	0.947	0.948	0.962	0.863
Interdependency	0.913	0.919	0.939	0.794
Top Management Support	0.955	0.956	0.967	0.881
Cultural Similarity	0.877	0.878	0.924	0.802
Goal Compatibility	0.938	0.941	0.956	0.845

Discriminant validity was assessed using the HTMT approach. The HTMT test has been identified as a rigorous test for discriminant validity and necessitate that the HTMT values must be significantly less than 1 (Henseler et al., 2015). Examining the HTMT values for the two competing models however revealed several HTMT ratios that were close to 1, indicating that discriminant validity might be a problem in this study.

Table 3a: HTMT results for model 1

	CI	COMMIT	II	INTDP	SI	Trust
CI						
COMMIT	0.898					
II	0.971	0.931				
INTDP	0.926	0.956	0.922			
SI	0.957	0.923	0.996	0.940		
Trust	0.906	0.965	0.928	0.982	0.938	

Table 3b: HTMT results for model 2

	CI	CUL	GOCOMP	II	SI	TMS
CI						
CUL	0.919					
GOCOMP	0.964	0.992				
II	0.971	0.888	0.956			
SI	0.957	0.897	0.954	0.996		
TMS	0.946	0.980	0.994	0.949	0.952	

Structural Model Results

Assessment of the structural model results was performed after first establishing the reliability and validity of the various constructs. The model fit was first assessed using the blindfolding analysis to assess the predictive relevance of the study. Hair et al. (2019) indicate that predictive relevance signifies the adequacy of a model to predict observed indicators of each latent constructs. Predictive relevance was signified by the assessment as Q^2 values. The Q^2 values for model 1 ranged between 0.732 to 0.897 whilst Q^2 values for model 2 ranged between 0.761 and 0.897, all of which are greater than zero, confirming predictive relevance of both models. The co-efficient of determination (R^2) for model 1 was 0.861 which suggests that about 86.1% of the variation in SCI was predicated by relationship drivers of SCI, which is a high level (Hair *et al*., 2019). On the other hand, the co-efficient of determination (R^2) for SCI for model 2 was 0.895, which suggests that about 89.5% of the variation in SCI was predicated by organizational drivers of SCI, which is slightly higher than the relationship drivers. Given that the predictive relevance and co-efficient of determination of SCI is slightly higher for the organizational drivers compared to the relationship drivers, it stands to suggests that the organizational drivers model (model 2) is a stronger predictor of the level of SCI. Thus, TMS, cultural similarity and goal compatibility better explain and predict the level of SCI than trust, commitment and interdependency.

The researchers further examined each hypothesized path to further gain insights into the drivers of SCI. To achieve this, the path coefficients and the t-values for the two research models were evaluated. The path coefficient shows the strength of the relationship between the constructs whereas the t-values show the significance of the path co-efficient. The results of the hypotheses tests are presented in Figure 2 and summarized in the Table 4 below.

Figure 2: Structural model results

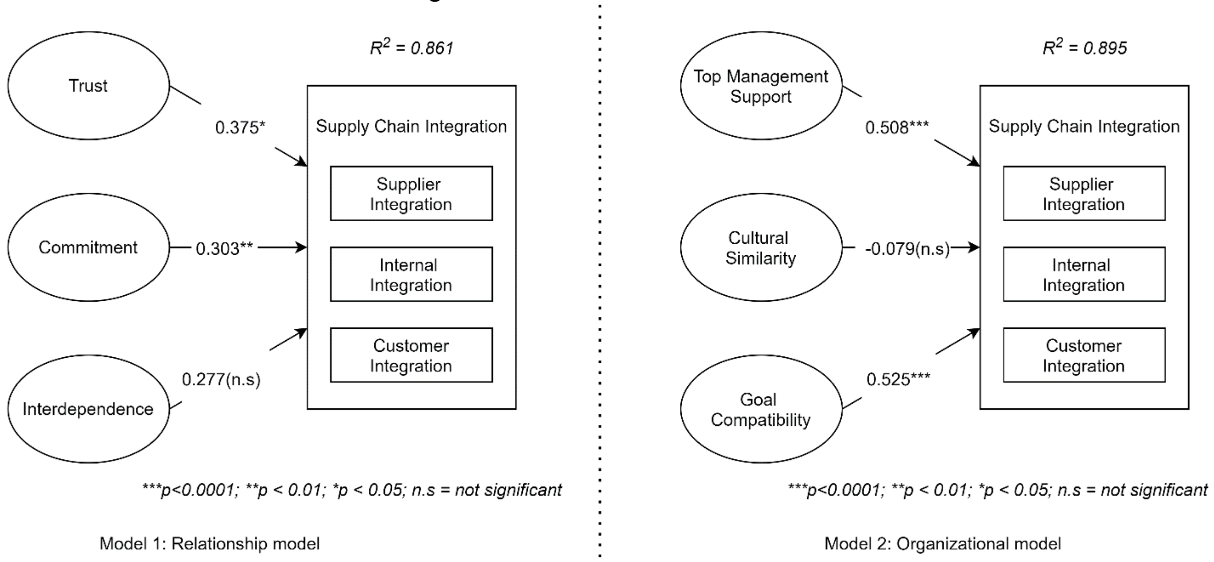


Table 4: Summary of model 1 hypotheses results

Hs	Hypothesized path	Beta	t value	p value	Decision
H1a	Trust → SCI	0.375	2.404	0.016	Supported
H2a	Commitment → SCI	0.303	2.626	0.009	Supported
H3a	Interdependency → SCI	0.277	1.805	0.071	Not Supported
H1b	Top Management Support → SCI	0.508	6.033	0.000	Supported
H2b	Cultural Similarity → SCI	-0.079	0.556	0.578	Not Supported
H3b	Goal Compatibility → SCI	0.525	4.875	0.000	Supported

From the bootstrapping analysis performed, it was observed that for both models, 2 out of 3 hypotheses were supported. For the relationship drivers of SCI (model 1), trust had a positive and significant effect on SCI thereby providing support for hypothesis 1a. Commitment had a positive and significant effect on SCI as hypothesized. However, interdependency was found to have positive but an insignificant impact on SCI, thereby not supporting the stated hypothesis. For the organizational drivers (model 2), TMS and goal compatibility were found to have positive and significant effects on SCI as hypothesized. One organizational driver, namely cultural similarity, was found to have a negative and an insignificant impact on SCI, thereby failing to support the stated hypothesis.

Discussions

The findings of the study revealed that trust had positive and significant effect on SCI (β = 0.375, t = 2.404, $p < 0.05$). This finding supports Simchi-levi and Simchi-levi (2007) assertion that trust is important to building cross-boundary integration and Vijayasarathy (2010) who found trust to have a positive and significant effect on supply integration. Also, the effect of commitment on SCI was positive and significant (β = 0.303, t = 2.626, $p < 0.05$). This finding is in line with Vijayasarathy (2010) study which found commitment to have a significant effect on supply integration and also Zhao et al. (2008) which found relationship commitment to be associated with customer integration. From the relational view perspective, commitment constitute vital relational norms for developing and maintaining relationships among supply chain partners such as SCI. On the contrary the last relationship driver, interdependency had a positive but an insignificant effect on SCI ($\beta = 0.277$, $t = 1.805$, $p > 0.05$). This finding fails to support the assertion of Seggie et al. (2006) that partner dependence boosts inter-firm system integration and Vijayasarathy (2010) findings that mutual dependence has a significant effect on supply integration. The relationship drivers largely had a positive and largely significant effect on SCI. This is not surprising particularly from an African context where mechanisms for enforcing contracts are not well developed, thereby necessitating the leveraging of relational factors such as trust and commitment to influence organizational initiatives such as SCI (Asamoah et al., 2021). Such findings demonstrate the need for firms particularly in Sub-Saharan Africa to build strong relational ties and social networks with key supply chain partners for the effective functioning of the supply chain (Asamoah et al., 2020).

Similarly, the organizational drivers revealed mixed outcomes on SCI. The effect of TMS on SCI was positive and significant ($\beta = 0.508$, $t = 6.033$, $p < 0.05$). This finding aligns with Xu et al. (2014) study which found top management support to have a significant effect on supplier and customer integration. Such finding also support the assertion that TMS is important in creating favourable climate and developing management styles which are vital to accomplishing the benefits of SCI implementation (Zhao et al., 2015). Additionally, the effect of goal compatibility on SCI was positive and significant ($\beta = 0.525$, $t = 4.875$, $p < 0.05$). Such finding stress on the need for organizations to integrate with supply chain partners that share the same or similar business goals as goal compatibility or congruence will help in achieving synergy which will result in mutual benefit for all partners. Lastly, the effect of cultural similarity on SCI was negative and not significant. This finding fails to support Rajaguru and Jekanyika (2012) finding that cultural compatibility has a positive impact on supply chain capabilities. This finding indicates that firms with varying cultures would still be able to achieve high levels of SCI.

Examining the relationship model and organizational factors model as competing theories reveal that organizational based model driven by the resource-based view was a stronger predictor of the level of SCI given its larger co-efficient of determination and higher levels of statistical support.

Conclusion

The study examined drivers of SCI using the competing theories and models approach. The relationship model, driven by the relational view, proposed that SCI was driven by trust, commitment and interdependency. On the other hand, the organizational model, driven by the RBV, proposed that SCI was driven by TMS, cultural similarity and goal compatibility. The study found that both relationship drivers and organizational drivers had largely positive and significant effects on SCI. Specifically, two relationship drivers (trust and commitment) had positive and significant effects on SCI. Similarly, two organizational drivers (TMS and goal compatibility) had positive and significant effects on SCI. Overall, the organizational drivers of SCI were slightly better at predicting the level of SCI than the relationship drivers.

The research has theoretical and practical implications. Theoretically, the study highlights the role of both relationship drivers and organizational drivers in understanding the level of SCI. Prior research has rarely examined how both the relationship and organizational drivers of supply chain activities influence SCI. This study thus examines the relationship between the drivers and SCI and confirms that both the relationship and organizational drivers are effective in predicting the level of SCI of firms. In addition to this, the study points to organizational drivers as slightly more important in predicting the level of SCI. The study also provides two models that can be used in predicting the effectiveness of SCI particularly in African context where contract enforcement mechanisms are not so effective.

Practically, the findings of the study stress on the need for firms to build relationships with their supply chain partners while ensuring the organizational setting is set to support SCI initiatives. Firms can leverage on the relationship drivers of trust and commitment to enhance their SCI implementations. Second, there is the need to ensure organizational drivers of TMS and goal compatibility, as it will serve as a valuable resource to achieve higher levels of SCI. The study is also important in providing guidance on how to achieve high levels of SCI within the Sub-Saharan African context. This is important as previous studies have observed that findings from other regions may not be supported within the unique Sub-Saharan African context (Asamoah et al, 2015; Asamoah et al., 2016; Agyei-Owusu et al., 2016).

The study has some limitations. First, the data used for the study was relatively small. Second, data utilized for the study was gathered from only Ghana, a Sub-Saharan region and as such the findings may not be generalizable to other contexts. Finally, discriminant validity could be a problem in this study as some HTMT values were observed to be closer to 1. Subsequent research can use larger sample size while others can focus on either manufacturing or service firms alone.

References

Agyei-Owusu, B., Asamoah, D., Andoh-Baidoo, F., & Akaribo, J. (2016). Impact of supply chain collaboration on logistics performance: evidence from a Sub-Saharan Nation’s petroleum downstream. *Decision Sciences Institute Meeting Proceedings*, Austin, Texas, USA.

Allred, C. R., Fawcett, S. E., Wallin, C., & Magnan, G. M. (2011). A dynamic collaboration capability as a source of competitive advantage. *Decision sciences*, *42*(1), 129-161.

Amit, R., & Shoemaker, P. (1993). Specialized assets and organizational rent. *Strategic Management Journal*, 14(1), 33–47.

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., Andoh-Baidoo, F. K., & Agyei-Owusu, B. (2015). Impact of ERP implementation on business process outcomes: A replication of a United States study in a Sub-Saharan African Nation. *AIS Transactions on Replication Research*, 1(1), 1-19.

Asamoah, D., Andoh-Baidoo, F. K., & Agyei-Owusu, B. (2016). Examining the relationships between supply chain integration, information sharing, and supply chain performance: a replication study. *Twenty-second Americas Conference on Information Systems*, San Diego.

Asamoah, D., Nuertey, D., Agyei-Owusu, B., & Acquah, I. N. (2021). Antecedents and outcomes of supply chain security practices: the role of organizational security culture and supply chain disruption occurrence. *International Journal of Quality & Reliability Management*.

Bacq, S., & Eddleston, K. A. (2018). A resource-based view of social entrepreneurship: How stewardship culture benefits scale of social impact. *Journal of Business Ethics*, *152*(3), 589-611.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.

Barney, J. B. (1986). Organizational culture: can it be a source of sustained competitive advantage? *Academy of Management Review*, 11(3), 656–665.

Boonstra, A. (2013), How do top managers support strategic information system projects and why do they sometimes withhold this support?. *International Journal of Project Management*, 31(4), 498-512.

Capaldo, A. (2014). Network governance: A cross-level study of social mechanisms, knowledge benefits, and strategic outcomes in joint-design alliances. *Industrial Marketing Management*, *43*(4), 685-703.

Capaldo, A., & Giannoccaro, I. (2015). How does trust affect performance in the supply chain? The moderating role of interdependence. *International journal of production Economics*, *166*, 36-49.

Cousins, P. D., & Menguc, B. (2006). The implications of socialization and integration in supply chain management. *Journal of operations management*, *24*(5), 604-620.

Danese, P. (2013). Supplier integration and company performance: A configurational view. *Omega*, 41(6), 1029-1041.

Dangayach, G. S., & Deshmukh, S. G. (2001). Manufacturing strategy: literature review and some issues. *International journal of operations & production management*. 21(7), 884–932.

Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of management review*, *23*(4), 660-679.

Eddleston, K. A., Kellermanns, F. W., & Sarathy, R. (2008). Resource configuration in family firms: Linking resources, strategic planning and technological opportunities to performance. *Journal of Management Studies*, *45*(1), 26-50.

Emden, Z., Calantone, R. J., & Droge, C. (2006). Collaborating for new product development: Selecting the partner with maximum potential to create value. *Journal of Product Innovation Management*, 23(4), 330–341.

Flynn, B. B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of operations management*, *28*(1), 58-71.

Ganesan, S., 1994. Determinants of long-term orientation in buyer–seller relationships. *Journal of Marketing*, 58 (2), 1–19.

Hair, J. F., Risher, J. J., Sarstedt, M. and Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2-24

Hamel, G., & Prahalad, C. K. (1996). Competing in the new economy: Managing out of bounds. *Strategic Management Journal*, 17(3), 237–242.

He, Y., & Lai, K. K. (2012). Supply chain integration and service oriented transformation: Evidence from Chinese equipment manufacturers. *International Journal of Production Economics*, *135*(2), 791-799.

Hendijani, R., & Saeidi Saei, R. (2020). Supply chain integration and firm performance: the moderating role of demand uncertainty. *Cogent Business & Management*, *7*(1), 1760477.

Huo, B. (2012). The impact of supply chain integration on company performance: an organizational capability perspective. *Supply Chain Management: An International Journal*, 17(6):596–610.

Huo, B., Qi, Y., Wang, Z., & Zhao, X. (2014). The impact of supply chain integration on firm performance: The moderating role of competitive strategy. *Supply Chain Management: An International Journal*, 369–384.

Jones, R. A., Jimmieson, N. L., & Griffiths, A. (2005). The impact of organizational culture and reshaping capabilities on change implementation success: The mediating role of readiness for change. *Journal of Management Studies*, 42(2), 359–384.

Koufteros, X. A., Rawski, G. E., & Rupak, R. (2010). Organizational integration for product development: the effects on glitches, on-time execution of engineering change orders, and market success. *Decision Sciences*, *41*(1), 49-80.

Koufteros, X., Vonderembse, M., & Jayaram, J. (2005). Internal and external integration for product development: the contingency effects of uncertainty, equivocality, and platform strategy. *Decision sciences*, *36*(1), 97-133.

Krause, D. R. (1999). The antecedents of buying firms' efforts to improve suppliers. *Journal of operations management*, *17*(2), 205-224.

Lee, B. C., Kim, P. S., Hong, K. S., & Lee, I. (2010). Evaluating antecedents and consequences of supply chain activities: an integrative perspective. *International Journal of Production Research*, *48*(3), 657-682.

Li, S., & Lin, B. (2006). Accessing information sharing and information quality in supply chain management. *Decision Support Systems*, 42(3), 1641–1656

Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013). Effects of supply chain integration and market orientation on firm performance: Evidence from China. *International Journal of Operations & Production Management*.

Lockström, M., Schadel, J., Harrison, N., Moser, R., & Malhotra, M. K. (2010). Antecedents to supplier integration in the automotive industry: a multiple-case study of foreign subsidiaries in China. *Journal of Operations Management*, *28*(3), 240-256.

Lu, D., Ding, Y., Asian, S., & Paul, S. K. (2018). From supply chain integration to operational performance: The moderating effect of market uncertainty. *Global Journal of Flexible Systems Management*, *19*(1), 3-20.

Mahapatra, S. K., Narasimhan, R., & Barbieri, P. (2010). Strategic interdependence, governance effectiveness and supplier performance: A dyadic case study investigation and theory development. *Journal of Operations Management*, *28*(6), 537-552.

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.

Mital, M., Chang, V., Choudhary, P., Papa, A., & Pani, A. K. (2018). Adoption of Internet of Things in India: A test of competing models using a structured equation modeling approach. *Technological Forecasting and Social Change*, *136*, 339-346.

Narasimhan, R., Swink, M., & Viswanathan, S. (2010). On decisions for integration implementation: An examination of complementarities between product-Process technology integration and supply chain integration. *Decision sciences*, *41*(2), 355-372.

Noordewier, T. G., John, G., & Nevin, J. R. (1990). Performance outcomes of purchasing arrangements in industrial buyer-vendor relationships. *Journal of marketing*, *54*(4), 80-93.

Panayides, P. M., & Lun, Y. V. (2009). The impact of trust on innovativeness and supply chain performance. *International journal of production Economics*, *122*(1), 35-46.

Rajaguru, R., & Matanda, M. J. (2013). Effects of inter-organizational compatibility on supply chain capabilities: Exploring the mediating role of inter-organizational information systems (IOIS) integration. *Industrial Marketing Management*, *42*(4), 620-632.

Rich, M. K. (2003). Requirements for successful marketing alliances. *Journal of Business & Industrial Marketing*, 18(4/5), 447–456

Samaddar, S., Nargundkar, S., & Daley, M. (2006). Inter-organizational information sharing: The role of supply network configuration and partner goal congruence. *European journal of operational research*, *174*(2), 744-765.

Schoenherr, T., & Swink, M. (2012). Revisiting the arcs of integration: Cross-validations and extensions. *Journal of operations management*, *30*(1-2), 99-115.

Seggie, S. H., Kim, D., & Cavusgil, S. T. (2006). Do supply chain IT alignment and supply chain interfirm system integration impact upon brand equity and firm performance?. *Journal of business research*, *59*(8), 887-895.

Simatupang, T. M., Wright, A. C., & Sridharan, R. (2002). The knowledge of coordination for supply chain integration. *Business process management journal,* 8(3), 289-309.

Swink, M., Narasimhan, R., & Wang, C. (2007). Managing beyond the factory walls: effects of four types of strategic integration on manufacturing plant performance. *Journal of operations management*, *25*(1), 148-164.

Tsanos, C. S., & Zografos, K. G. (2016). The effects of behavioural supply chain relationship antecedents on integration and performance. *Supply Chain Management: An International Journal.*

Van der Vaart, T., & Van Donk, D. P. (2008). A critical review of survey-based research in supply chain integration. *International journal of production economics*, *111*(1), 42-55.

Vijayasarathy, L. R. (2010). Supply integration: an investigation of its multi-dimensionality and relational antecedents. *International Journal of Production Economics*, *124*(2), 489-505.

Wathne, K. H., & Heide, J. B. (2000). Opportunism in interfirm relationships: Forms, outcomes, and solutions. *Journal of marketing*, *64*(4), 36-51.

Wong, A., Tjosvold, D., & Zhang, P. (2005). Supply chain relationships for customer satisfaction in China: Interdependence and cooperative goals. *Asia Pacific Journal of Management*, *22*(2), 179-199.

Wong, C. Y., Boon-Itt, S., & Wong, C. W. (2011). The contingency effects of environmental uncertainty on the relationship between supply chain integration and operational performance. *Journal of Operations management*, *29*(6), 604-615.

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*.

Yang, C. C., Marlow, P. B., & Lu, C. S. (2009). Assessing resources, logistics service capabilities, innovation capabilities and the performance of container shipping services in Taiwan. *International Journal of Production Economics*, *122*(1), 4-20.

Yuan, Y., Feng, B., Lai, F., & Collins, B. J. (2018). The role of trust, commitment, and learning orientation on logistic service effectiveness. *Journal of Business Research*, *93*, 37-50.

Yuen, K. F., & Van Thai, V. (2017). The influence of supply chain integration on operational performance: A comparison between product and service supply chains. *The International Journal of Logistics Management*.

Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization science*, *9*(2), 141-159.

Zhao, G., Feng, T., & Wang, D. (2015). Is more supply chain integration always beneficial to financial performance?. *Industrial Marketing Management*, *45*, 162-172.

Zhao, X., Huo, B., Selen, W., & Yeung, J. H. Y. (2011). The impact of internal integration and relationship commitment on external integration. *Journal of operations management*, *29*(1-2), 17-32.

DECISION SCIENCES INSTITUTE
On risk-return profiling of Peer-to-Peer platforms

Janis Eismonts
Sneakypeer
janis@sneakypeer.com

Michel Verlaine
ICN Business School
Michel.verlaine@icn-artem.com

ABSTRACT

This paper aims to develop a risk-return profiling approach for Peer-to-Peer platforms. We develop a top-down approach to risk budget different subcomponents of Peer-to-Peer platforms. As the risk distributions of the Peer-to-Peer platforms are typically skewed with extreme risks, we use coherent measures of risk and shortfall measures that can be aggregated consistently across different risks. This approach can be used by decision makers to choose optimal portfolio combinations of Peer-to-Peer loan portfolios and analyze the risk profiles of platforms. We apply the method to Lending Club data and we extract the loss distributions for the different risk buckets of loan portfolios, which enables us to evaluate expected returns as well as the expected losses and the expected shortfalls. Our method can be used as a risk management and portfolio optimization tool for any decision maker who can fix his confidence level that implies a certain aggregate risk budget.

KEYWORDS: Risk Management, FinTech, Portfolio selection, Coherent aggregation

JEL CLASSIFICATION: G11, D81, C11

INTRODUCTION

Recent financial innovations and the development of digital technologies have considerably changed the landscape and the business models of the finance industry. Hence, the development of Fintech, defined by the Financial Stability Board (2017) as “technologically enabled financial innovation that could result in new business models, applications, processes or products with an associated material effect on financial markets and institutions and the provision of financial services”, might have major impacts on the sector of credit institutions, a process analyzed by Stulz (2019). This has led, however, to concerns about financial stability if this sector grows. The sector of fintech credit poses specific regulatory and risk management challenges (CGFS-FSB (2017)). Although, there is no internationally agreed definition, it broadly includes all credit activity implemented through electronic platforms not operated by commercial banks. It encompasses activity implemented through borrower and lender matching platforms. In different jurisdictions those platforms might take different denominations like “peer-to-peer (P2P) lenders”, “loan-based crowd-funders” or “marketplace lenders” and be regulated in different ways. Those platforms also differ in the way they use their own balance sheets, provide guarantees, work with loan originators and have institutional investors. Those platforms that belong to the so-called alternative credit market are also increasingly becoming a new asset class for some investors. In the US, for instance, approximately 60% of P2P funding is financed by institutional investors and the activity

is now referred to as “marketplace lending”. Recent trends include institutional investors funding bulks of credits with eventually some securitization process involved (US Department of the Treasury (2018)). The different types of platforms with their different business models, however, are likely to hide different types of risks. Whereas financial risks are typically measured with risk measures adapted to measure market, credit and liquidity risk, the different business model risks, have to be analyzed through agency mechanisms stemming from asymmetries of information between different parties. Ideally, the risk inherent in such a platform should be quantified and aggregated with the investment risk implicit in the structure’s portfolio. The third type of risk stems from operational risk. As the degree of digitalization of the platforms is much higher than for other institutions the main operational risk is given by cyber risks, a serious threat for such structures. To assess the aggregate risk of the different types of platforms, some kind of aggregate risk measure needs to be developed. The problem, however, is that the risk distributions of the different components differ. The task is thus to aggregate risks for different types of risk distributions. Such a process supposes that risk measures are coherent (Artzner *et al*. (1999)). This is a problem of coherent aggregation of risk measures analyzed by Kalkbrener, M. (2005), Dorfleitner, G. and Buch, A. (2008), Fischer, T. (2003). The typical risk measure that respects all coherency principles and is adapted to the different risk distributions is expected shortfall (Acerbi, C. and Tasche, D. (2002)).
The modelling of credit risks also poses particular challenges. For an overview of consumer credit risks (Livshits (2015)). Recently, the operations research (OR) literature has become interested in the modelling of P2P platforms’ loss distributions. Most studies focus on the two largest P2P platforms in the US, namely Prosper and Lending Club. The credit supply process implicit in P2P lending, however, differs from standard retail banking. Even though P2P platforms screen potential borrowers against their own criteria, once the loan is listed on the platform the investors bear the risks. This is in contrast with bank lending where banks bear the risk and the get the return. It is important to highlight that P2P platforms have different business models that might also entail different kinds of risks. The business model of the platform is thus also of foremost importance.
The P2P platforms have recently attracted, not only retail investors, but also institutional investors. For instance, Balyuk and Davydenko (2018) indicate that for the Prosper platform, 90% of loans were provided through institutional channels. They also suggest that active strategies may yield more than passive institutional strategies. Thus, a framework to analyze the risk-return profiles of P2P portfolios and/or platforms is of foremost interest.

The literature on P2P credit scoring is divided between two potential approaches. The operational research approach and the finance approach. The first focusses on P2P loan scoring methods. Recent approaches are developed in Kim *et al*. (2019), Verbraken *et al*. (2014), Serrano-Cinca and Gutierrez-Nieto (2016), Guo *et al*. (2016). The finance perspective also focuses on how investors adapt the platforms’ operations, business models and available information (Miller (2015), Balyuk and Davydenko (2018), Jagtiani and Lemieux (2017), Vallee and Zheng (2019)). Apart from factors linked to creditworthiness, the finance literature also focuses on softer information. The softer information, however, is more costly to verify than harder information (Liberty and Peterson (2019)).
Recently, Fitzpatrick and Mues (2021) investigated the combination of different approaches. Their study is based on Lending Club data and they analyze loan level predictors based on a combination of loans, borrowers, credit risk and text derived characteristics.
As the marketplace lending platforms are a new type of emerging institutions, not much is known about potential risk management solutions. Claessens *et al*. (2018) also argue that recent experiences lead to fundamental challenges for users, regulators and investors.

- What is the degree of risk transparency and disclosure of the platforms and do investors and borrowers really understand them?

- Are the interest rates and fees of the platform in line with the potential risk and expectations?
- What is the underlying degree of complexity and what kinds of risks does it hide?
- Are the operational risk management, notably data- and cyber-security systems adequate?

The aim of this paper is to provide a solution to those challenges by analyzing those platforms' ecosystems and develop an integrated risk management approach.
The objective is to measure the risk for each component so to be able to diagnose the specific risks of each component for a respective platform as well the aggregate risk, namely total aggregate risk of investment, knowing that the risk might be allocated to different substructures. Such an aggregation procedure, however, is not straightforward and depends on the risk measures that are used in the process. We use coherent risk and shortfall measures that are consistent with a coherent aggregation process. This means that we can set risk budget targets that can be consistently aggregated making sure that the total risk is in line with the target risk.
In section 1, we describe the process to measure and aggregate risk in a coherent way. Section 2 presents the portfolio optimization approach to determine the risk-return allocation in the FinTech structure. In section 3, we use a factor model approach to model the loss distribution of underlying portfolios. Finally, section 4 presents the empirical implementation of the method.

1. Financial Risk and Coherent Aggregation

Capital allocation is obviously a major topic for financial institutions in order to determine the global risk buffer and to benchmark different business units to ensure that value added is in line with the risks incurred. One of the major challenges, however, rests with the aggregation and decomposition of risk capital across different business lines. The BIS defines economic capital as "the methods or practices that allow financial institutions to attribute capital to cover economic effects of risk-taking activities". The ability of economic capital models to adequately reflect business-line operating practices and therefore provide appropriate incentives to business units is also an issue. Finally, our major challenge, here, is the development of any simple risk measure to capture adequately all the complex aspects of Fintech platform risks.
Economic capital is the amount of capital a financial institution needs to absorb unexpected losses over a certain time horizon at a given confidence level. Expected losses are typically accounted for in the pricing of a financial institution's products and loan loss provisions. It is thus unexpected losses that require economic capital. Issues with the implementation of economic capital allocation arise at different levels. Credit portfolio management aims at improving the risk-adjusted return profiles of credit portfolios via credit risk transfer transactions and control of the loan approval process. Loan's marginal contributions to the portfolio's economic capital and risk-adjusted performance are the major criteria. The economic capital can then be used to price the credit products. Risk-adjusted performance is measured with RAROC. In order for economic capital to be totally integrated in the internal decision process, economic capital needs to be integrated into the objective functions of the decision makers.
Typically, financial institutions' relative performance is measured with risk-adjusted performance measures based on economic capital. The most common risk-adjusted performance measures are RAROC (Risk-adjusted Return on Capital). Whereas RAROC measures performance compared to risk. In order to evaluate RAROC, a hurdle rate, which reflects the financial intitutions's cost of capital, as well as a confidence level to evaluate economic capital have to be set. Economic capital plays an important role in the hypothetical capital allocation for the budgeting process. This process is an important component of the strategic planning and the

target setting. Strategic planning involves setting the risk appetite, whereas target setting involves, for instance, deciding on an external rating. The economic capital allocation process may depend on the governance structure of the financial institution, in our case the business model of Fintech credit platform. In a standard financial institution, after economic capital is allocated, the business unit's managers are supposed to manage risk so that it does not exceed economic capital. Depending on how centralized the management of the bank is, internal allocation of capital at the business unit is delegated to the business managers. In the case of Fintech credit platforms, the process might be diluted as there is no centralized evaluation of risks and economic capital with risk limits.
The most widely used risk measures are Value at Risk (VaR) and Expected Shortfall (ES). VaR is the most intuitive measure and it is the easiest to communicate. However, as it is not sub-additive whenever risks are not elliptically distributed, it can cause problems in terms of internal capital allocation and limit settings. ES is less easily interpreted, but is generally sub-additive and can be used to set risk limits consistent with the aggregate target risk. In order for those risk measures to be calculated, a confidence interval has to be fixed. The latter is often function of an external rating target and the confidence interval is interpreted as the economic capital necessary to prevent an erosion of the capital buffer. The relationship between external ratings and PD's is not necessarily very stable and the targeting of an external rating might be difficult to realize. Other considerations such as risk allocation to different business lines to measure profitability might call for lower confidence levels. Incidentally, the confidence interval might make portfolios look relatively better or worse depending on the distribution of extreme risks.

As already alluded to, a main issue rests with the aggregation and decomposition of risk measures. Decomposition refers to the fact that within a portfolio, risk needs to be decomposed to analyse the risk inherent to each position. Decomposition is important to determine capital allocation, limit setting, pricing of products and risk-adjusted performance measurement. Aggregation properties are important to determine the financial institutions's overall economic capital from market, credit, operational and other risks.
In order to aggregate risks, financial institutions have to categorize the sources of risk into risk units. They are typically categorized along two dimensions: the economic nature of the risks, namely credit, market and operational risks, and the organizational structure of the financial institution. Even though classification along organizational lines such as business lines is straightforward, the disentangling of market, credit and operational risk can be quite difficult. In the case of fintech trading platforms, given the eventual complexity of the Business Models, the risk inherent in the organizational structure might be more difficult to evaluate, a topic further discussed in the next section.
The main difficulty with aggregation however rests with the elaboration of the unit of risk account, also known as common risk currency. Typically, units of risk account depend on three characteristics. First, the risk metric should be in line with the risk metrics used to measure individual risk components and, here, sub-additivity of the risk metric is a major issue. Second, when distributions of risks do not share the same distributional assumptions, the specification of confidence levels is likely to impact the risk aggregation process. Finally, the determination of the time horizon for the aggregation is not straightforward, as the different risk components (market and credit notably) are measured in different units of time. Typically, the time horizon is fixed at one year.
Different aggregation methodologies are used by financial institutions. The easiest way to aggregate the risks is to use the sum of the risks. This is a conservative measure of economic capital as it is not taking into account potential diversification effects and might lead to lower performance on economic capital. Some institutions thus implement a slightly more advanced approach by fixing a diversification percentage coefficient. The approach presumes that the adding of risks leads to a predetermined diversification effect, the latter being difficult to specify

in practice. Aggregation can also be implemented on the basis of a variance-covariance matrix, even though this approach bears on the assumption that risks are multivariate normally distributed. In order to circumvent this problem, the most consistent approach is to model common risk drivers across all portfolios and aggregate the risks with a copula, which allows for risk aggregation with marginal risk components that are not normally distributed.
Our approach, here, is to evaluate the risk capital that is implicit in the activities of the FinTech Credit platforms activities. We thus have to aggregate the different risk capital viewed as some kind of risk budget for different structures involved in the FinTech Credit Business Model. The aggregate risk capital can then be compared to the aggregate effective capital held by the FinTech credit Business Model, to determine the implicit probability of default, which can be used to map it on a potential rating. In order to analyse that question more formally we need to focus more in detail on the properties of the risk measures used.
(Artzner et al., 1999) derived the axioms that a well-behaved risk measure should respect. A so-called coherent risk measure $\rho(X)$ is a real valued function defined on the space of real-valued random variables, which satisfies the following axioms:

$$\forall c \epsilon \Re, \rho(X+c) = \rho(X) - c \tag{1}$$

$$\forall X, Y, \rho(X+Y) \leq \rho(X) + \rho(Y) \tag{2}$$

$$\forall \lambda \geq 0, \rho(\lambda X) = \lambda \rho(X) \tag{3}$$

$$X \geq 0, \ \ \rho(X) \leq 0 \tag{4}$$

All those axioms have a specific interpretation. Translation invariance (1), also called cash-invariance, implies that the risk can be interpreted in terms of capital requirements, a property we need to compare risk capital to the effective capital held by the FinTech credit Business Model. If an investor puts the amount $\rho(X)$ of cash in the structure, the structure becomes riskless under a certain probability. It thus indicates the amount of risk capital the structure should have, given the risk measure used ρ. Positivity (4), of course, just implies that a higher payoff profile has less risk. A measure that satisfies both axioms is called a monetary measure of risk (Föllmer & Scheid, 2008). Positive homogeneity (3) means that an increase in all the portfolio positions by λ also leads to a linear increase in the risk involved. This axiom has actually been questioned on the grounds of liquidity risk. In our case, it implies that the increase in volume of the FinTech structure does not increase the risk more than proportionally. Finally, sub-additivity (2) which is probably the most important axiom implies that the risks of a portfolio are less than the sum of individual risks. A well-behaved risk measure should thus model diversification effects. This property enables us to aggregate the different risk capital values across the whole FinTech structure in a consistent way.

This last property can also be used from an active management perspective to decentralize the task of managing the risks arising from a collection of different components of the FinTech Credit Structure. Separate risk limits can potentially be set for the different components making sure that the risk of the aggregate FinTech Credit Platform is bounded by the sum of individual risk limits, which of course would be conditioned by the default probability. That default probability can then be linked to certain target rating. In many risk management problems, the following VAR measure is used.

$$VaR_{\alpha} = inf\{m \in \mathcal{R} | P[X+m < 0 \leq \alpha]\} \tag{5}$$

As alluded to earlier, VaR is not sub-additive when tail risk is high which is quite likely the case for some risks in the FinTech credit structure.

Expected shortfall given by

$$ES_\lambda = \frac{1}{\lambda}\int_0^\lambda VaR_\alpha(X)d\alpha \tag{6}$$

is a coherent risk measure and can be used as an effective risk measure to set decentralized risk limits. ES also has other interesting properties in terms of manager control and robustness. As highlighted by (Basak & Shapiro, 2001), risk limits in terms of VAR may lead to portfolio concentrations and increases of extreme risks. ES measures can also be deduced from robust decision making in the face of uncertainty (Föllmer & Schied, 2004). The development of coherent risk measures has also led to research on coherent capital allocation which is of particular interest here.
Coherent capital allocation has been a research topic since the seminal work of Artzner et al. (1999) and is important to consistently allocate risk capital to the different subunits of the FinTech structure. The first major contribution to the issue of coherent economic capital allocation was provided by Denault (2001). He analyses how adding different business units or sub-portfolios leads to diversification effects, thus the sum of the risks of the subcomponents is larger than the risk of the sum of the subcomponents. We want to analyse the total risk capital of a FinTech structure taking into account the diversification effects and dividing the risk capital among the sub-units.

Kalkbrener (2005) develops an axiomatic approach and shows that sub-additivity and positive homogeneity of risk measures is of foremost importance. The suggested axiomatization presumes that the risk capital allocated to a sub-unit *i* depends exclusively on the risk of sub-unit *i* and the risk of the whole FinTech structure, not on the decomposition of the other sub-units. The axiomatization of Kalkbrener (2005) is based on the assumption that capital allocated to sub-unit SU_i only depends on SU_i and the aggregate FinTech Structure *S*, but not on the decomposition of the rest of the portfolio S$-SU_i = \sum_{i\neq j} SU_i$. The capital allocation rule is thus represented by a function $\Phi(SU_i, S)$. If we consider an aggregate risk measure $\rho(S)$, fixed by a decision maker for instance, then Φ is a capital allocation with respect to $\rho(S)$ whenever $\Phi(S,S) = \rho(S)$, hence when the capital allocated to the aggregate portfolio equals the risk capital of the aggregate FinTech structure.

Three axioms are important for coherent capital allocations across the FinTech structure. First, linear aggregation makes sure that the sum of the risk capital of the sub-units equals the risk capital of the aggregate FinTech structure. Second, as mentioned earlier there are some diversification effects. Third, continuity ensures that small sub-unit adjustments have a limited impact of the risk capital of the aggregate structure. Those axioms uniquely characterize capital allocations. Also, for a given risk measure ρ there exists a capital allocation $\Phi_\rho$that satisfies the linear aggregation and diversification axioms if and only if ρ is sub-additive and positively homogeneous, that is ρ satisfies:

$$\rho\left(SU_i + SU_j\right) \leq \rho(SU_i) + \rho(SU_j) \tag{7}$$

$$\rho(\lambda SU_i) = \lambda\rho(SU_i), \forall\lambda \geq 0 \tag{8}$$

The existence of directional derivatives of ρ at the aggregate level S is a necessary and sufficient condition for Φ_ρ to be continuous at *S*.

The following *iff* conditions can be used to set up the coherent risk capital process across the Fintech credit structure.

- If there exists a linear, diversifying capital allocation Φ_ρ with respect to ρ, then ρ is positively homogeneous and sub-additive.
- If ρ is positively homogeneous and sub-additive then Φ_ρ is a linear, diversifying capital allocation with respect to ρ.

We suggest now to integrate shortfall measures in an optimizing framework. This is an approach pioneered by Bertsimas *et al*. (2004). Instead of maximizing the return for a given risk, the return is mapped on potential risk factors and the distance between the return and the expected shortfall is then minimized. The allocation should be optimal given an aggregate shortfall measure that could be determined by an investor or the decision maker of a FinTech credit structure. The aggregate measure can then be used to evaluate the aggregate capital of the FinTech Structure to the aggregate risk capital for a given targeted rating. We first develop the general mathematical framework before elaborating on the application to the specific FinTech Structure.

2. General Portfolio optimization with shortfall measures

Let's now consider a vector of expected returns $E(R)$ that are potentially linked to the risk profile of a subset of borrowers, with mean

$$E(R) = \mu \tag{9}$$

and a vector of portfolio weights **ω**. As we will concentrate on risk measurement and rating of a FinTech structure, the mean return will be fixed at a target return rate for a portfolio of borrowers.

$$E[R^T\omega] = R_p \tag{10}$$

Depending on the respective DM, however, this value could be fixed as expected loss if they are provisioned and we have a risk management perspective.
As mentioned earlier, we could now minimize over **ω** the (1-α) confidence level

$$VaR_\alpha(\omega) := \mu^T\omega - q_\alpha(R^T\omega), \forall\alpha \in (0,1) \tag{11}$$

where $q_\alpha(R^T\omega)$ is the α-quantile of the distribution of the portfolio return $R^T\omega$. The $q_\alpha(R^T\omega)$ would then be fixed to be consistent with a certain probability of default linked to the parameter α. A natural approach of portfolio optimization would thus consist in minimizing VaR_αas a function of portfolio weights. From an operational perspective the FinTech structure should thus be such that it minimizes the VaR_α for a given return of the underlying return on the sub-portfolios of borrowers.

Issues with sub-additivity rather lead us to use shortfall to impose limits on deviations of losses with respect to the expected return. We thus suggest to minimize the shortfall at the risk level α and as we reason in terms of losses:

$$s_\alpha(\omega) := \mu^T\omega - E[L^T\omega | L^T\omega \geq q_\alpha(L^T\omega)], \forall\alpha \in (0,1) \tag{12}$$

$s_\alpha(\omega)$ basically, measures the expected losses below the expected return that might occur whenever the portfolio drops below the α-quantile. It measures how much is lost on average when things go really bad. Given those results we are led to focus on shortfall minimization. In order to analyze the risk capital of the FinTech structure we need a risk budgeting process.
The risk measure originally introduced by (Artzner et al, 1999) was not expected shortfall but Tail Conditional Expectation

$$TCE_\alpha(\omega) = -E[L^T\omega | L^T\omega \geq q_\alpha(L^T\omega)] \quad (13)$$

Which is a coherent risk measure.

As shortfall is to be used in portfolio optimization it integrates the expected return and shortfall can be expressed as the following way:

$$s_\alpha(\omega) = \mu^T\omega + TCE_\alpha \quad (14)$$

Due to this mean-adjustment, expected shortfall violates translation invariance and positivity.
Shortfall has other useful mathematical properties. Most notably, an important issue for portfolio optimization is that it is convex in the portfolio weights ω which is not necessarily the case with VaR.
Moreover:

- $s_\alpha(\omega) \geq 0\ \forall\omega\ and\ \alpha \in (0,1)$. $s_\alpha(\omega)$ equals 0 for some ω and α if and only $L^T\omega$ is constant with probability 1.
- $s_\alpha(\omega)$ is positively homogeneous: $s_\alpha(\lambda\omega) = \lambda s_\alpha(\omega), \forall\lambda \geq 0$
- In case the density of losses is continuous, we get $\nabla_\omega s_\alpha(\omega) = \mu - E[L | L^T\omega \geq q_\alpha(L^T\omega)]$

Those properties are useful to create a risk budgeting process that can be used to analyse the risks of the FinTech structure but also guide an operational risk management process.

If we can simulate the loss distribution of a reference portfolio of loans, shortfall can be estimated quite easily. Consider a simulated sample of N losses of a portfolio of loans ω. The losses can then be ranked from the lowest indexed by 1 to the highest indexed by *N*. $l_i(\omega)$ and the losses can be sorted in increasing order:

$$l_1(\omega) \leq l_2(\omega) \leq \ldots \leq l_N(\omega) \quad (15)$$

Consider $K = [\alpha N]$, the number of losses above a certain level, the non-parametric estimator of $s_\alpha(\omega)$ can be represented thus:

$$s_\alpha(\omega) = \omega^T\mu - \frac{1}{K}\sum_{j=1}^{K} l_{(j)}(\omega) \quad (16)$$

From an operational risk management viewpoint, the following mean-shortfall optimization can be solved to determine the optimal allocation of risk within a FinTech structure.
Minimize $s_\alpha(\omega)$
subject to $\omega^T\mu = r_p\ and\ e^T\omega = 1$
where e is a column vector of ones.

Bertsimas et al. (2004) show that the minimum α-shortfall frontier is convex. In presence of a riskless asset the minimum shortfall is defined from:
Minimize $s_\alpha(\omega)$
Subject to $\omega^T\mu + (1 - e^T\omega)r_f = r_p$
(Tasche, 2000) has shown that the minimum shortfall frontier in the $\left(r_p, s_\alpha(\omega)\right)$ space, with $r_p \geq r_f$, is a ray starting from the point $\left(r_f, 0\right)$. and passing through the particular point $\left(r_p^*, s_\alpha^*(\omega)\right)$ with $r_p^* \geq r_f$.
The optimal solution satisfies

$$\mu_j - r_f = \beta_{j,\alpha}(\omega_\alpha)\left(r_p - r_f\right), j = 1, \dots, n \tag{17}$$

With

$$\beta_{j,\alpha}(\omega) = \frac{1}{s_\alpha(\omega)} \times \frac{\partial s_\alpha(\omega)}{\partial \omega_j} \tag{18}$$

$$= \frac{\mu_j - E\left(L_j | \omega^T L \geq q_\alpha(L^T\omega)\right)}{\omega^T\mu \quad - E(\omega^T L | \omega^T L \geq q_\alpha(L^T\omega))} \tag{19}$$

$\beta_{j,\alpha}$is called the shortfall beta and can be used to measure the relative change in shortfall when varying and allocating loans across the FinTech structure. Note that the following characteristic that will be useful for the risk budgeting process holds.

$$\sum_{j=1}^{n} \omega_j \, \beta_{j,\alpha}(\omega) = 1 \tag{20}$$

This is a very useful property as it gives a decomposition of a portfolio shortfall into individual contributions of loans or substructures of the FinTech structure. As mentioned earlier it can be used to measure the risk capital of any financial structure but also as a tool where the management can fix a shortfall target and given Euler's Theorem applies to the measure, to deduce:

$$s_\alpha(\omega) = \sum_{j=1}^{n} \omega_j \times \frac{\partial s_\alpha(\omega)}{\partial \omega_j} \tag{21}$$

More exactly:

$$\mu \quad (\omega) - E(\omega^T L | \omega^T L \geq q_\alpha(L^T\omega)) = \sum_{j=1}^{n} \omega_j \times \left[\mu_j - E\left(L_j | \omega^T L \geq q_\alpha(L^T\omega)\right)\right] \tag{22}$$

Different approaches can now be used to evaluate or simulate the loss distributions.

3. The factor model approach

This approach underlies portfolio models based on a structural approach where default occurs when some value of a latent process falls below a limit. The main advantage is that it reduces the dimensionality of the dependence problem for large portfolios. A latent factor drives the default process. When the value of the latent factor is below a threshold value, the borrower defaults. The latent factors are assumed to depend on a common factor as well as an idiosyncratic term specific to each borrower and uncorrelated to the common factor. Within the standard approach, the factors are assumed normally distributed. This is the so-called Gaussian Copula model. In general, the factors are assumed to be normally distributed and are scaled to have unit variance and zero expectation. In case, of heavy tail dependence of the risks, however, the model can be adapted to take into account stronger tail distributions of the factors. This then leads to the so-called student *t*-copula. As the market standard for pricing credit risk is the Vasicek (2002) asymptotic single factor model, we will consider the model in more detail here and eventually adapt it later to a more complex environment.

It is assumed that the dynamic of an obligor n's wealth is given by a continuous-time geometric Brownian motion:

$$\frac{dS_{n,t}}{S_{n,t}} = \mu_n dt + \sigma_n d\varepsilon_{n,t} \tag{23}$$

where μ_n is the expected return, σ_n the instantaneous volatility and $\varepsilon_{n,t\,t}$ the standard Brownian motion.

It is now assumed that default occurs when the Brownian motion falls below a limit L. The probability of default, given the value of the Brownian motion at time t, is given by:

$$p_{n,t,T} = P\left[S_{n,T} < L \middle| S_{n,t}\right] \tag{24}$$

By applying Itô's lemma we can derive the value of the Brownian motion at time T as a function of its current value:

$$S_{n,T} = S_{n,t}\, exp\left\{\left(\mu_n - \frac{{\sigma_n}^2}{2}\right)(T-t) + \sigma_n\sqrt{T-t}\,d\varepsilon_{n,t,T}\right\} \tag{25}$$

where $d\varepsilon_{n,t,T}$ is given by:

$$d\varepsilon_{n,t,T} = \frac{\varepsilon_{n,T}-\varepsilon_{n,t}}{\sqrt{T-t}}. \tag{26}$$

Default then amounts to $d\varepsilon_{n,t,T} < K_{n,t,T}$ where

$$K_{n,t,T} = \frac{lnL_n - ln\,S_{n,t} - \left(\mu_n - \frac{\sigma_n^2}{2}\right)(T-t)}{\sigma_n\sqrt{T-t}} \quad (27)$$

The calculations are similar to the ones implemented to derive the B&S formula and the probability of default can hence be expressed by : $p_{n,t,T} = \Phi(K_{n,t,T})$ where Φ is the distribution function of standard normal random variable. By simulating different paths of the Brownian motion, we can determine the default probability for different borrowers *n*. If we can find similar types of reference borrowers, the limit K can also be calibrated by inverting the phi function $\phi^{-1}(p_n) = K$. There might, however, exist some default correlation that we have to take into account in the shock process. This is typically done with a factor model.

In consistency with the model above, consider a portfolio of N obligors where each obligor defaults as soon as the relevant latent factor $S_{n,t}$ drops below L. In order to derive a closed form for the portfolio default rate, namely the fraction of defaulted credits in the portfolio at time t, Vasicek (2002) made a few assumptions. The portfolio default rate will be determined by the individual default rates as well as the default correlations. Those correlations will affect the random variables $d\varepsilon_{n,t}$ and create common effects accross borrowers. Vasicek assumes that the correlation coefficient between any pair of random variables is the same for any two obligors, hence:

$$corr(d\varepsilon_{n,t}, d\varepsilon_{m,t}) = \rho_{n,m,t} = \rho_t \forall n \neq m \quad (28)$$

A common correlation coefficient then partly drives all the obligor's latent factor processes. The idea is that there exists a random systemic factor that impacts the latent factor of all the obligors. The random variables triggering default can now be written:

$$d\varepsilon_{n,t} = \sqrt{\rho}M_t + \sqrt{1-\rho}z_{n,t} \quad (29)$$

Default is determined by two factors: M_t which is a common or systematic factor and $z_{n,t}$which is an idiosyncratic factor. M_t generates the default dependency as it affects all obligors in the same way. This model is equivalent to the Gaussian copula. The default probability at time *t* can now be evaluated by conditioning on the common factor M_t.

$$p_n(M_t) = P[d\varepsilon_{n,t} < K_{n,t}|M_t] \quad (30)$$

$$= P\left[z_{n,t} < \frac{K_{n,t} - \sqrt{\rho_t}M_t}{\sqrt{1-\rho_t}} | M_t\right]$$
$$= \Phi\left(\frac{K_{n,t} - \sqrt{\rho_t}M_t}{\sqrt{1-\rho_t}}\right)$$

Conditionally on M_t the default probabilities are independent. It is assumed that we know the individual default probabilities and we can calculate $K_{n,t} = \Phi^{-1}(p_{n,t})$. Note that as $(p_{n,t})$ is a function of market spreads, we could in principle infer the default probabilities from the spreads in market data for similar types of borrowers. If we further assume that the default probabilities are equivalent for all obligors, we get the following expression for conditional default probability:

$$p(M_t) = \Phi\left(\frac{\Phi^{-1}(p_t) - \sqrt{\rho_t}M_t}{\sqrt{1-\rho_t}}\right) \tag{31}$$

Now, consider the fraction of losses on the portfolio at time *t*: FL_t. The unconditional distribution function of a portfolio characterized by a default probability *p* and correlation *ρ* is

$$F(FL^*|p_t \ ,\rho_t) = P(FL_t \leq FL^*) \tag{32}$$

If we assume that the number of obligors tends to infinity, then due to the conditional independence assumption the fraction of losses FL_t converges to the individual default probability $p(M_t)$. The unconditional distribution function of the fraction of losses can hence be expressed thus:

$$F(FL^*|p_t \ ,\rho_t) = P\left(FL_t \leq FL^*\right) \tag{33}$$
$$= P(p(M_t) \leq FL^*)$$
$$= P\left[\Phi\left(\frac{\Phi^{-1}(p_t) - \sqrt{\rho_t}M_t}{\sqrt{1-\rho_t}}\right) \leq FL^*\right]$$
$$= P\left[M_t \geq \frac{\Phi^{-1}\left(p_t - \sqrt{1-\rho_t}\right)\Phi^{-1}(FL^*)}{\sqrt{\rho_t}}\right]$$
$$= 1 - \Phi\left(\frac{\Phi^{-1}(p_t) - \sqrt{1-\rho_t}\Phi^{-1}(FL^*)}{\sqrt{\rho_t}}\right)$$
$$= \Phi\left(\frac{\sqrt{1-\rho_t}\Phi^{-1}(FL^*) - \Phi^{-1}(p_t)}{\sqrt{\rho_t}}\right)$$

By changing the value of the fraction of losses FL^* for different steps and evaluating the cumulative probability, we can calculate the cumulative distribution function of the fraction of losses also called default rate. Deriving the latter distribution gives us the density of the default rate that can be used to calculate the loss distribution of the portfolio.

4. Empirical implementation

The density can also be calculated more directly by deriving the cumulative distribution over the fraction of losses. The above-mentioned relationship can then be expressed as the following density for the fraction of losses *FL*.

$$\left(\varphi(FL||p_t\ \ ,\rho_t\,\right) = \sqrt{\frac{1-\rho}{\rho}}\,exp\left(\frac{1}{2}\left\{[\Phi^{-1}(FL)^2] - \left[\frac{\sqrt{1-\rho}\,\Phi^{-1}(FL)-\Phi^{-1}(p_t)}{\sqrt{\rho}}\right]^2\right\}\right) \tag{34}$$

The density function can then be estimated by maximum likelihood. The maximum likelihood function is a function that depends on the probability of default and the coefficient of correlation where we will use the fraction of losses as an independent variable. The density $\left(\varphi(FL||p_t\ \ ,\rho_t\,\right)$ and the likelihood parameters are those that maximise the following product

$$\prod_{t=1}^{T}\left(\varphi(FL||p_t\ \ ,\rho_t\,\right) \tag{35}$$

and it is typically transformed by applying so to maximize the log-likelihood function

$$\sum_{t=1}^{T} ln\left(\varphi(FL||p_t\ \ ,\rho_t\,\right) \tag{36}$$

Which more explicitly gives

$$\sum_{t=1}^{T} ln\sqrt{\frac{1-\rho}{\rho}} + \frac{1}{2}\left\{[\Phi^{-1}(FL)^2] - \left[\frac{\sqrt{1-\rho}\,\Phi^{-1}(FL)-\Phi^{-1}(p_t)}{\sqrt{\rho}}\right]^2\right\} \tag{37}$$

For standard credit portfolios, the time horizon is typically one year but for the fintech platform the credits are shorter so the method can be applied on shorter horizons such as one month or even less. The method then amounts to calculating the logarithm of the probability density of the fraction

of losses for each observation on the fraction of losses. For the first step trial values for the probability of default and the coefficient of correlation are typically fixed. Basel guidelines for similar portfolios might be a good starting point. The solver or iteration can then estimate the values of the probability of default end the coefficient of correlation that maximise the log likelihood function which is the logarithm of the density for each observation. Those parameters can then be fed back into the formulas to determine the loss distribution.
We implement the method with data from Lending Club using a Taylor-made R code. We first estimate the probability of default unconditionally from the loan datasets with different riskiness. We then use the maximum likelihood approach to estimate the parameter of correlation ρ. Those parameters are then put into the Vasicek formula which provides probabilities over fractions of losses. We can then simulate different levels of fractions of losses for the different loan portfolios of different risk categories from A (lowest risk) to G (highest risk). From the loss distributions, we can estimate the expected loss, the shortfall measure $s_\alpha(\omega)$ for a given confidence interval α. The appendix presents the Figures with the properties of the Lending Club data as well as the estimated Vasicek distributions for each relevant loan category. We can now calculate the global target shortfall from investor risk profiles and determine the optimal portfolio weight using

$$\mu \quad (\omega) - E(\omega^T L | \omega^T L \geq q_\alpha(L^T \omega)) = \sum_{j=1}^{n} \omega_j \times \left[\mu_j - E\left(L_j | \omega^T L \geq q_\alpha(L^T \omega)\right)\right] \tag{38}$$

which provides the weight of each subcomponent to reach the target shortfall as elicited from a potential decision maker.

The table below indicates the estimated expected return and shortfall measures per risk bucket of Lending Club for a cutoff level α of 0.95. The weight1 column indicates the weights estimated using the above-mentioned equation.

	Expected_Return	shortfall	weight1	weight2
A	7.11	0.76	0.11	0.07
B	10.66	0.89	0.13	0.09
C	14.08	0.97	0.14	0.12
D	17.81	1.03	0.15	0.14
E	21.23	1.07	0.15	0.17
F	25.01	1.12	0.16	0.2
G	27.71	1.15	0.16	0.22

We also estimate the weights of the portfolio by evaluation the ratio of expected return over Expected shortfall. This is formally a risk adjusted performance measure (RAPM). By dividing each RAPM by the sum of RAPMs for all buckets we get the weight 2.

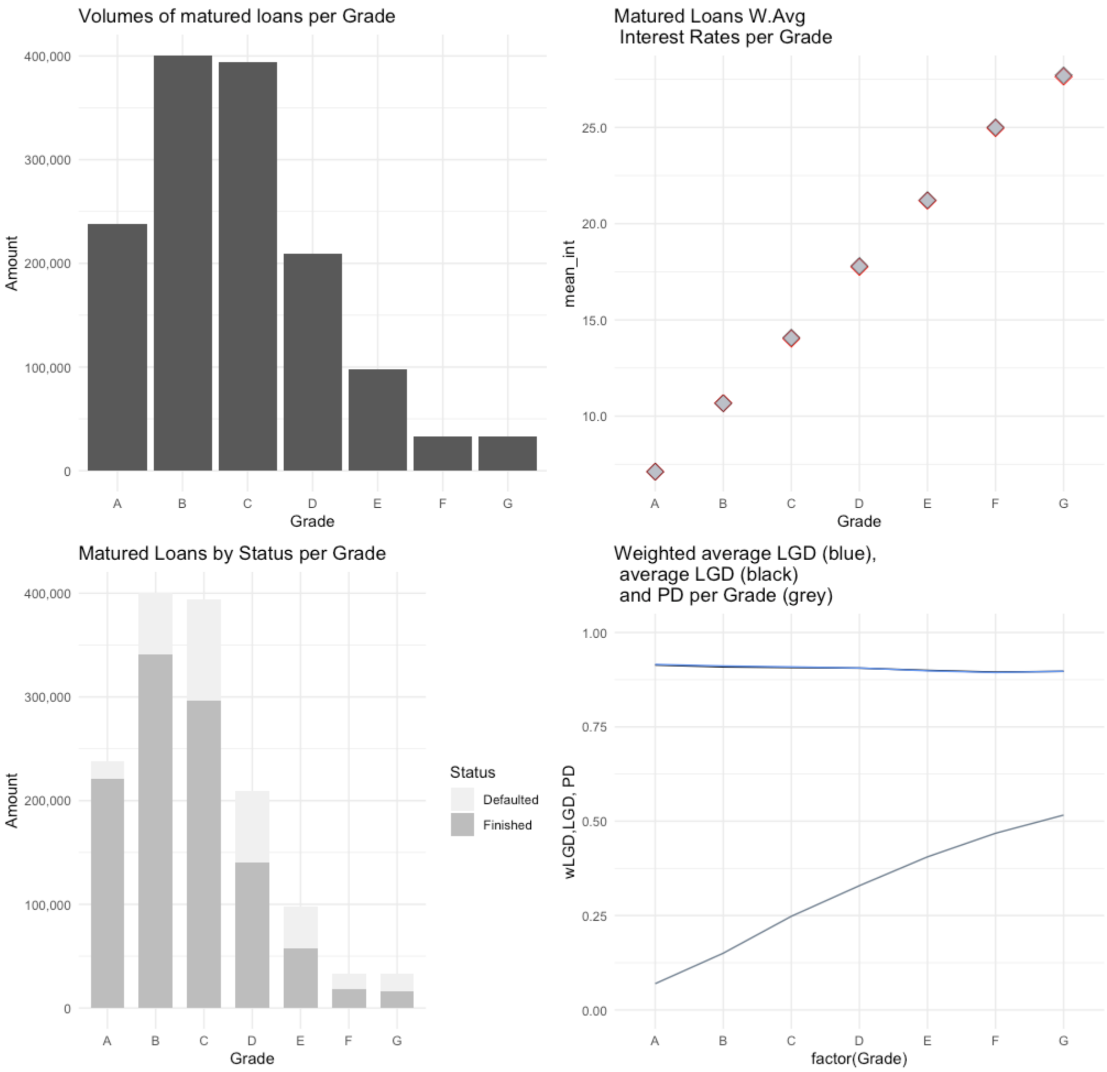
DISCUSSION AND CONCLUSIONS

This paper develops a portfolio and risk management approach for new emerging digital P2P platforms. We analyze the risk return profiles of different subunits of the P2P platform. We start by characterizing the risk measurement approach that is adapted to solve the coherent aggregation problem in a P2P platform. Then a portfolio approach based on shortfall measures is developed. Then the standard factor model approach with correlated Brownian motions enables us to characterize the loss distributions of the different portfolios. Using a Taylor-made R code we then estimate the different types of loss distributions from which we infer expected losses, expected shortfalls as well returns. This information can then be used in our portfolio optimization approach by any decision maker interested in risk management as well as investment. Further research will focus on implementation with different shortfall confidence intervals elicited from different risk profiles of decision makers.

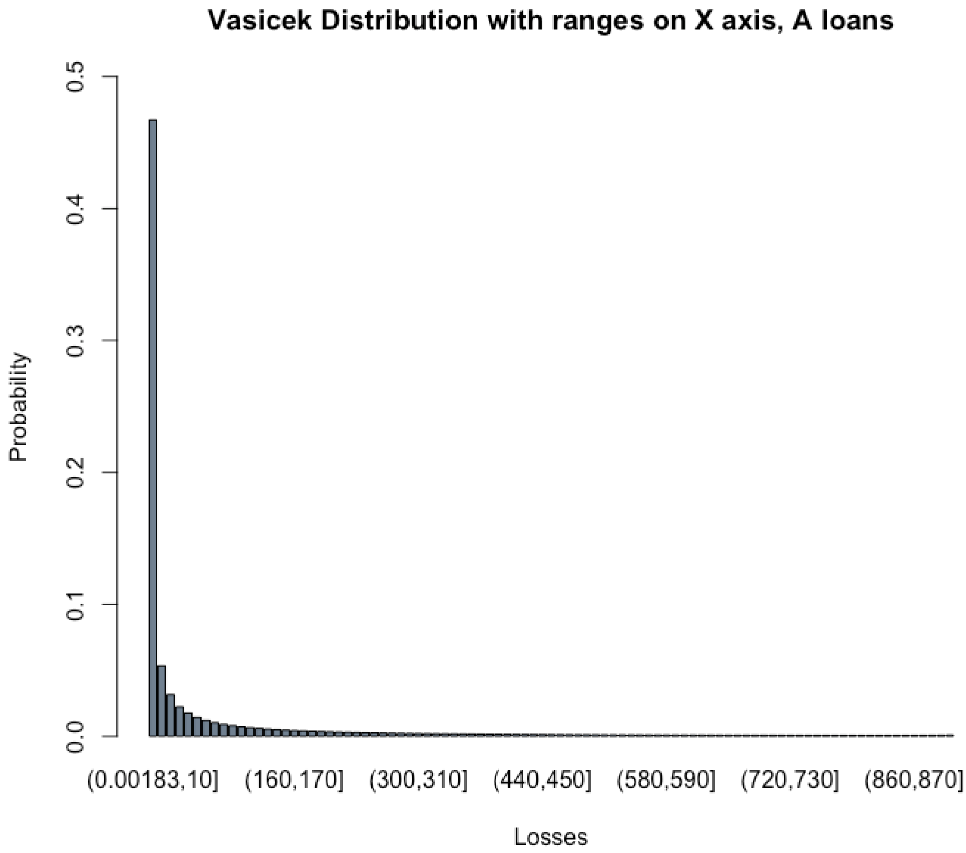
APPENDIX

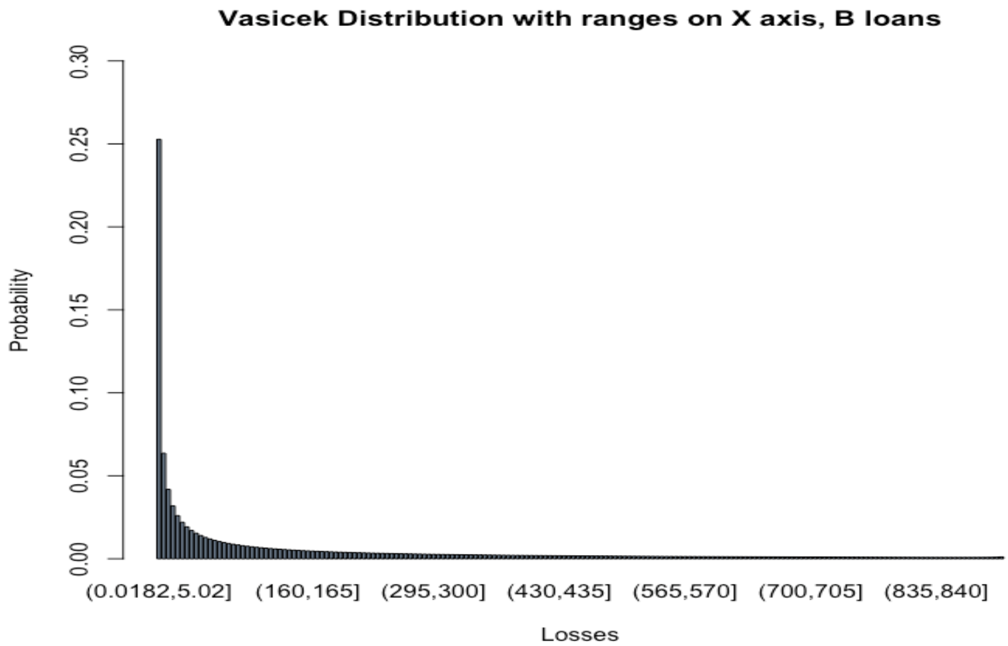
The appendix presents the collection of Figures evaluated from Lending Club data. We are grateful for research support and assistance from Alberts Vinklers and Emils Arcimovics from EcoFin Research for the coding in R.

On risk-return profiling of Peer-to-Peer platforms

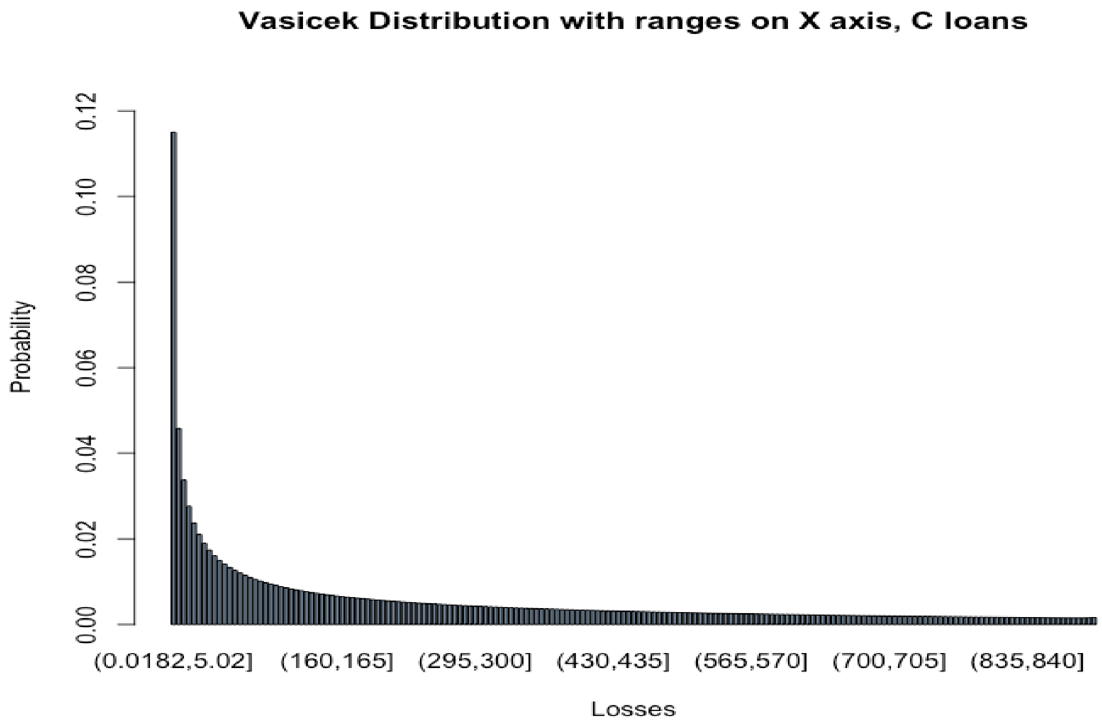


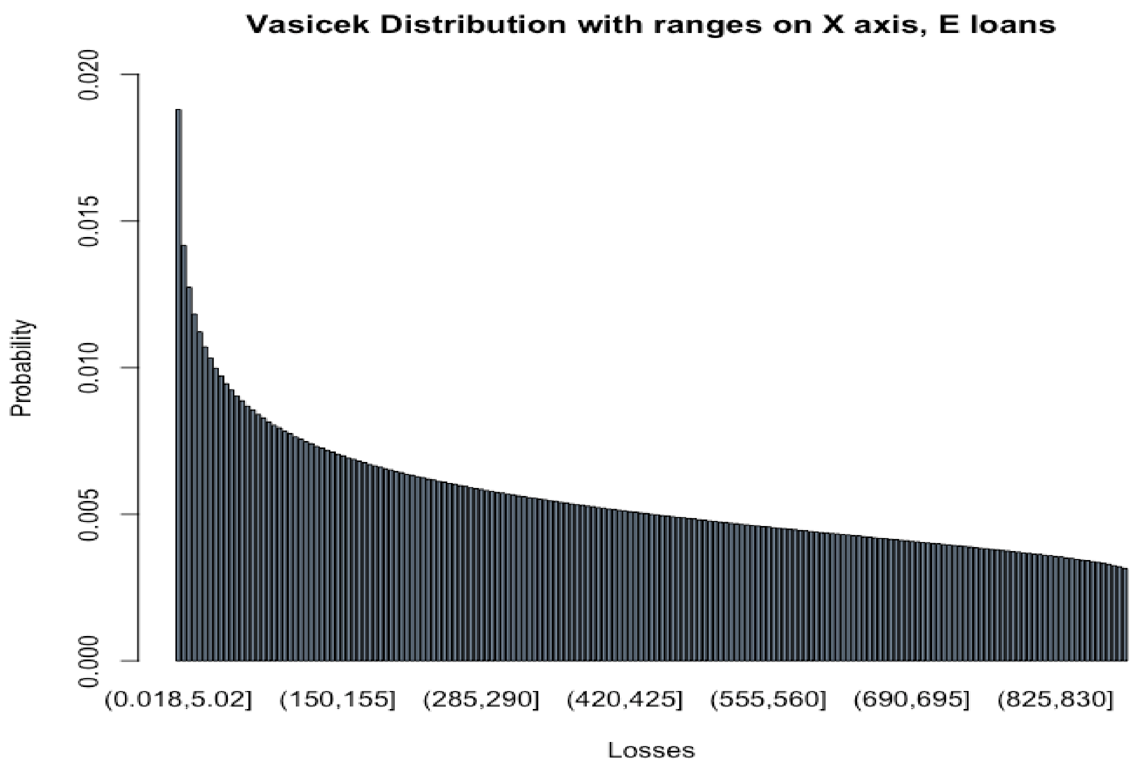
On risk-return profiling of Peer-to-Peer platforms

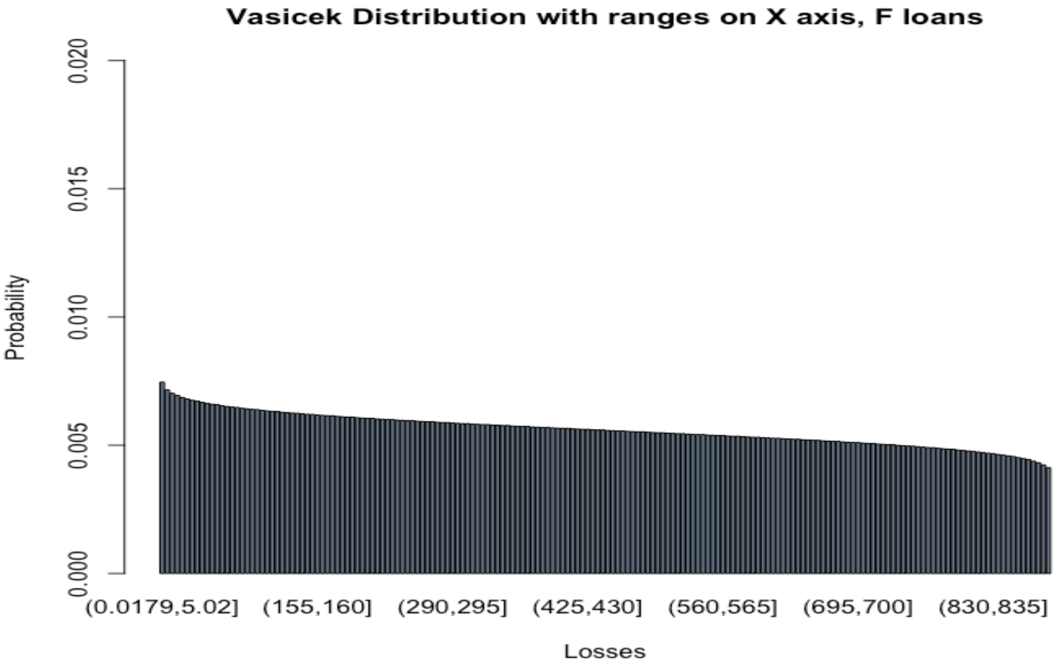


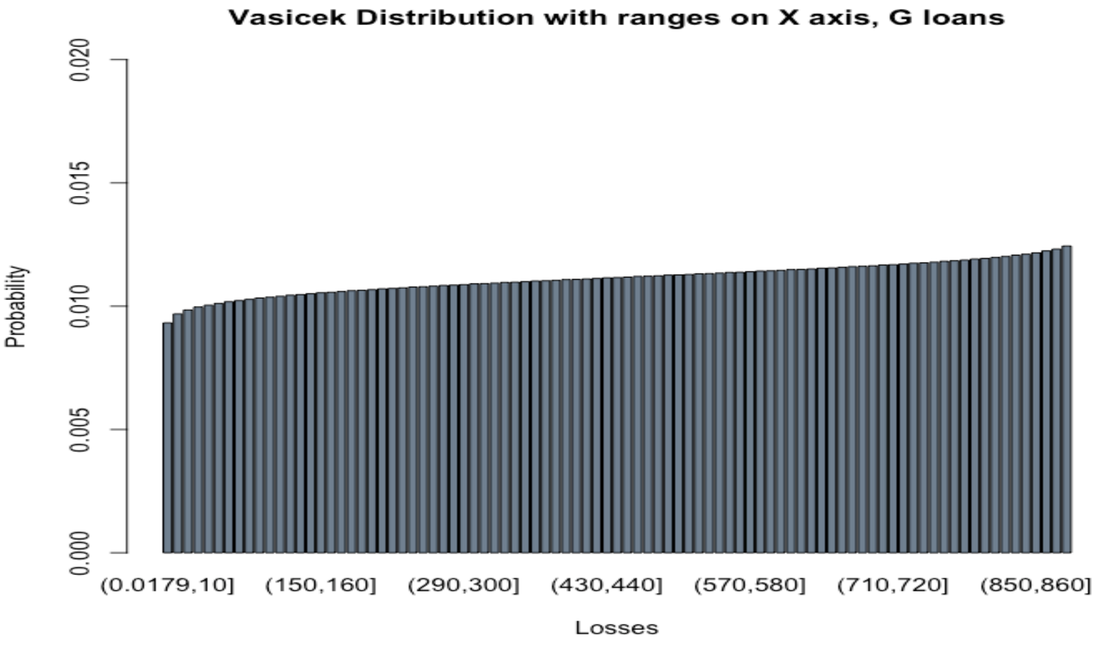


On risk-return profiling of Peer-to-Peer platforms






Vasicek Distribution with ranges on X axis, F loans
Probability
0.000
0.005
0.010
0.015
0.020
(0.0179,5.02]
(155,160]
(290,295]
(425,430]
(560,565]
(695,700]
(830,835]
Losses


Vasicek Distribution with ranges on X axis, G loans
Probability
0.000
0.005
0.010
0.015
0.020
(0.0179,10]
(150,160]
(290,300]
(430,440]
(570,580]
(710,720]
(850,860]
Losses

REFERENCES

Acerbi, C. & Tasche, D. (2002) "On the coherence of expected shortfall", *Journal of Banking and Finance*, Vol. 26, p. 1487-1503.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath D. (1999) "Coherent Measures of Risk", *Mathematical Finance*, Vol. 9, p. 203-228.

Balyuk, T. & Davydenko, S.A. (2018) "Reintermediation in Fintech: Evidence from online lending", *SSRN Working Paper.*

Basak, S. & Shapiro, A. (2001) "Value at Risk based risk management: Optimal policies and asset prices", *Review of Financial Studies*, p. 317-405.

Basel Committee on Banking Supervision (2020) "Sound management of risks related to money laundering and financing of terrorism", Guidelines.

Begley, T. & Purnanandam, A. (2017) "Design of financial securities: Empirical evidence from private-label RMBS deals", *Review of Financial Studies* 30, 120-161.

Bertsimas, D., Lauprete, G. & Samarov, A. "Shortfall as a risk measure: properties, optimization and applications", *Journal of Economic Dynamics and Control* 28, 1353-1381.

Cambridge Center for Risk Studies (2016) "Managing Cyber Insurance Accumulation Risk"
Claessens, St., Frost, J., Turner, G. & Zhu, F. (2018) "Fintech credit markets around the world: size, drivers and policy issues", *BIS Quaterly Review.*

Committee on the Global Financial System and Financial Stability Board (2017) "FinTech credit: market structure, business models and financial stability implications", *CGFS papers May*.

DeMarzo, P. & Duffie, D. "A liquidity-based model of security design", *Econometrica* 67, 65-99.

Denault, M. (2001) "Coherent allocation of Risk Capital", *Journal of Risk*, 4(1), p. 1-34.

Dorfleitner, G. & Buch, A. (2008) "Coherent risk measures, coherent capital allocations and the gradient allocation principle", *Insurance:Mathematics and Economics*, Vol. 42, p. 235-242.
EIOPA (2018) "Understanding Cyber Insurance – A Structured Dialogue with Insurance Companies", Publications office of the EU.

Fischer, T. (2003) "Risk Capital allocation by coherent risk measures based on one-sided moments", *Insurance: Mathematics and Economics*, Vol. 32, P. 135-146.

Föllmer, H. & Schied, A. (2008) "Convex and Coherent Risk Measures", *Encyclopedia of Quantitative Finance, John Wiley and Sons.*

Föllmer, H. & Schied, A. (2004) *Stochastic Finance: An Introduction in Discrete Time*, Gruyter Studies in Mathematices 27.

Fritzpatrick, T. & Mues, C. (2021) “How can lenders prosper? Comparing machine learning approaches to identify profitable peer-to-peer loan investments”, *European Journal of Operational Research,* 294(20, 711-722.

Garcia, R., Renault, E. & Tsafack, G. (2007) “Proper Conditioning for Coherent VaR in Portfolio Management”, *Management Science*, Vol. 53.

G-7 Cyber Expert Group (2017) *Fundamental Elements for Effective Assessment of Cybersecurity in the Financial Sector*, October. Washington DC: G-7
.
Guo, Y., Zhou, W., Luo, C., Liu, C. & Xiong, H. (2016) “Instance-based credit risk assessment for investment decisions in P2P lending”, *European Journal of Operational Research*, 249(2), 417-426.

Jiagtani, J. & Lemieux, C. (2017) “Fintech lending: Financial inclusion, risk pricing, and laternative information”, *Technical Report*, Federal Reserve Bank of Philadelphia
.
Kalkbrener, M. (2005) “An Axiomatic Approach to Capital Allocation”, Mathematical Finance, Vol. 15, p. 425-437.
Kim, A., Yang, Y., Lessman, S., Ma, T., Sung, M. & Johnson, J. (2019) “Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting”, *European Journal of Operational Research*, 1-18.

Koijen, R. & Yogo, M. (2016) “Shadow Insurance”, *Econometrica*, Vol. 84, 1265-1287.

Leland, H. & Pyle, D. (1977) “Informational asymmetries, financial structure and financial intermediation”, *Journal of Finance* 32, 371-387.

Lewis, J. (2018) *Economic Impact of Cybercrime—No Slowing Down.* Santa Clara McAfee.

Liberti, J.M. & Petersen, M.A. (2019) “Information: Hard and Soft”, *The Review of Corporate Finance Studies*, 8(1), 1-41.

Miller, S. (2015) “Information and default in consumer credit markets: Evidence from a natural experiment”, *Journal of Financial Intermediation* 24(1), 45-70.

Milne, A. & Parboreeah, P. (2016) *The business models and economics of Peer-to-Peer lending,* Euro Credit Res Inst.

Omarini, E. (2018) “Peer-to-Peer Lending: Business Model Analysis and the Platform Dilemma”, *International Journal of Finance, Economics and Trade*.
Serrano-Cinca, C. & Guttierez-Nieto, B. (2016) “The use of profit scoring as an alternative to credit scoring systems in pee-to-peer (P2P) lending”, *Decision Support Systems* 89, 113-122.

Stulz, R. (2019) “FinTech, BigTech and the future of Banks”, *NBER Working Paper Series.*
US Department of the Treasury (2018) *A financial system that creates economic opportunities: nonbank financials, fintech, and innovation*, July.

Valleee, B. & Zheng, Y. (2019) “Marketplace lending: A new banking paradigm?”, *The Review of Financial Studies,* 32(5), 1939-1982.

Vasicek, O. (2002) “Loan Portfolio Value”, Risk, 15, December, 160-162.
Verbraken, T., Bravo, C., Weber, R. & Baesens, B. (2014) “Development and application of consumer credit scoring models using profit-based classification measures”, 238(2), 505-513.

DECISION SCIENCES INSTITUTE

Text Mining for Pension De-Risking Analysis

Limin Zhang
North Dakota State University
Email: limin.zhang@ndsu.edu

Ruilin Tian
North Dakota State University
Email: ruilin.tian@ndsu.edu

Jeffrey (Jun) Chen
North Dakota State University
Email: jun.chen1@ndsu.edu

ABSTRACT

In the recent 20 years, sponsors of defined benefit (DB) pension plan have been facing severe underfunding challenges and have used a variety of methods to reduce pension-related risks such as investment risk, interest rate risk, credit risk, and longevity risk. In this study, we extract company filings between 1993 to 2018 from the SEC EDGAR database to identify different "de-risking" strategies that US-based firms have used. A combination of text mining, machine learning, and natural language processing methods has been applied to the textual data for automated classification of de-risking strategies. The primary contributions of this research are (1) the design of a multi-phase methodology that identifies and extracts information from large amount of textual data; (2) the development of a comprehensive database for pension de-risking activities of U.S. firms; and (3) an empirical analysis based on the pension de-risking database that provides valuable insights to companies with DB plans, pensioners, and practitioners in the de-risking markets.

KEYWORDS: Text mining, machine learning, natural language processing, classification, supervised learning, pension de-risking, SEC EDGAR

DECISION SCIENCES INSTITUTE

Supporting Services, Hedonic Tendencies, and E-impulse Buying on Travel Booking Websites

Kalyani Ankem
Northern Kentucky University
Email: ankemk1@nku.edu

Rupesh Agrawal
Northern Kentucky University
Email: agrawalr1@nku.edu

Vishal Uppala
North Dakota State University
Email: vishal.uppala@ndsu.edu

ABSTRACT

We present research on the effects of supporting services on web browsing and e-impulse buying. Hedonic rather than utilitarian browsing has a greater possibility of leading to impulse buying. We study two services offered by businesses to support customers' decision-making online: customer reviews and product recommendations. We analyze survey data using PLS-SEM to present findings. Literal review has a positive effect on utilitarian browsing (p<.001). However, utilitarian browsing has a negative effect on e-impulse buying (p<.05). Hedonic browsing has a positive effect on e-impulse buying (p<.05). Also, recommendation quality has a positive effect on e-impulse buying (p<.01).

KEYWORDS: Impulse buying, Hedonic browsing, Utilitarian browsing, Customer reviews, Recommendations

INTRODUCTION

Impulse buying is big business. Impulse buying is when customers make a decision to purchase something right before they purchase it. In other words, consumers have a sudden urge to buy something immediately when they see it. These are unplanned purchases which often involve consumers' emotions. As such, impulse buying has the potential to serve as a form of retail therapy by lifting consumers' spirits. Consumers spend a substantial amount on unplanned purchases. Businesses can place triggers to leverage the impulse buying phenomenon.
In this research, we are interested in how customers' perceptions of reviews and recommendations relate to their web browsing and e-impulse buying behavior. Web browsing includes both utilitarian and hedonic forms of browsing. We want to analyze the impact of recommendations on e-impulse buying. We also want to examine how customer reviews affect utilitarian browsing and hedonic browsing and whether utilitarian browsing and hedonic browsing lead to e-impulse buying.

HEDONIC WEB BROWSING AND E-IMPULSE BUYING

Sales promotion stimuli such as buy-one-get-one are known to facilitate online impulse buying (Lo et al., 2016). These sales promotion stimuli can generate both utilitarian and hedonic benefits for customers. Lo et al. (2016) categorize the utilitarian and hedonic benefits of sales promotions and rank the effectiveness of sales promotion stimuli on online impulse buying. Lo et al. do not examine the relationship between utilitarian benefits, hedonic benefits, and online impulse buying.

Other researchers show that hedonic browsing in online shopping leads to impulse buying. Chan et al. (2017) conducted a systematic review on impulse buying which supports the positive effects of hedonic shopping motivation and hedonic consumption needs on impulse buying. Kim and Eastin (2011), included in Chan et al.'s systematic review, examined the effect of hedonic shopping motivation on consumer online shopping behavior. According to Kim and Eastin, hedonic shopping motivation is a robust predictor of impulse buying. In Park et al.'s (2012) research on online apparel shopping, three factors related to apparel shopping had different effects on utilitarian versus hedonic browsing. Additionally, two of these factors, variety of selection and sensory attributes, had direct effects on impulse buying (Park et al., 2012). When analyzing the effects of utilitarian and hedonic browsing, Park et al. found that only hedonic browsing had a positive effect on impulse buying. Utilitarian browsing had a negative effect on impulse buying. Liao et al. (2016) illustrate similar relationships. According to these researchers, hedonic products generate pleasure which positively influences the urge to buy impulsively. In related research, Sela and Berger (2012) focus on one factor, the number of product attributes, to find its effect on people's choice of hedonic and utilitarian options. They show that product "attribute numerosity benefits hedonic more than utilitarian options" (p. 942). In their research, neither impulse buying nor purchase intention was investigated.

SUPPORTING SERVICES

According to Cenfetelli et al. (2008, p. 161),

> With the continued growth of business-to-consumer (B2C) e-business, online vendors are providing an increasing array of services that support and enhance their core products or services. For example, Amazon.com does not just sell books; it also enhances that core product with automated product recommendations, "wish list" tracking, order status updates, customer reviews, and many other valuable supporting services.

Cenfetelli et al. study the functionalities offered by such supporting services. They call these supporting services functionalities (SSF). We have chosen to study the supporting services instead and their contribution to the customer's decision to purchase online. We include two widely used, noninteractive supporting services: online customer reviews and product recommendations.

Reviews

Customers often consider feedback, when available, before making purchase decisions. Hence, customer reviews can affect consumer decisions online. Valence (Elwalda & Lü, 2016), level of detail (Jiménez & Mendoza, 2013), quantity (Elwalda & Lu, 2016), quality (Hong et al., 2018), and length (Godes & Mayzlin, 2004) are some dimensions of online customer reviews that have been analyzed for their effect on online consumer purchase behavior. By far, the impact of positive and negative reviews on consumer behavior and sales has received the most attention.

Few researchers have studied the effect of customer reviews on utilitarian versus hedonic choices. Kronrod and Danziger (2013) found that customer reviews containing more figurative language led to hedonic consumption. Hazari et al. (2017) examined hedonic and utilitarian use of user-generated content on online shopping websites to find how hedonic and utilitarian use affect purchase intention. Impulse buying is not addressed in their research. Hazari, et al. found that utilitarian “use of UGC [user-generated content] has a greater impact than hedonic use on intention to purchase” (p. 576).

Recommendations

Recommender systems or recommendation agents (RAs) offer recommendations to customers based on their interests or preferences. These recommendations support customers while they are shopping online. Xiao and Benbasat (2007) found that use and characteristics of recommendation agents influence consumer decision-making processes and outcomes. Characteristics of recommendation agents covered in the study are RA type, input, process, and output. Consumer decision-making outcomes in this study are consumers’ choice and decision quality. Hostler et al. (2011) are among the few researchers to study the influence of recommendation agents on impulse purchasing online. In Hostler et al.’s research, recommender agents increased unplanned purchasing online.

Some researchers have focused on social presence in recommendation agents. Choi et al. (2011, p. 129) found that while “greater social presence increases both reuse intention and trust in the recommender systems,” “the influence of social presence on reuse intention with respect to utilitarian products is less than that with respect to hedonic products.” The target variable in this research is recommender-system reuse intention. The hypotheses do not include impulse buying or purchase intention. The target variable in Wang et al.’s (2016) study is trust. They investigated the effects of rational appeals (explanation facilities) and social appeals (avatar interfaces) made by recommendation agents on cognition- and affect-based trust. This study, too, does not include impulse buying or purchase intention.

Other aspects of recommendations have been addressed in research. Roudposhti et al. (2018) illustrate how recommendation quality affects customer purchase intention. Recommendation novelty, diversity, and accuracy are shown to have a significant effect on recommendation quality.

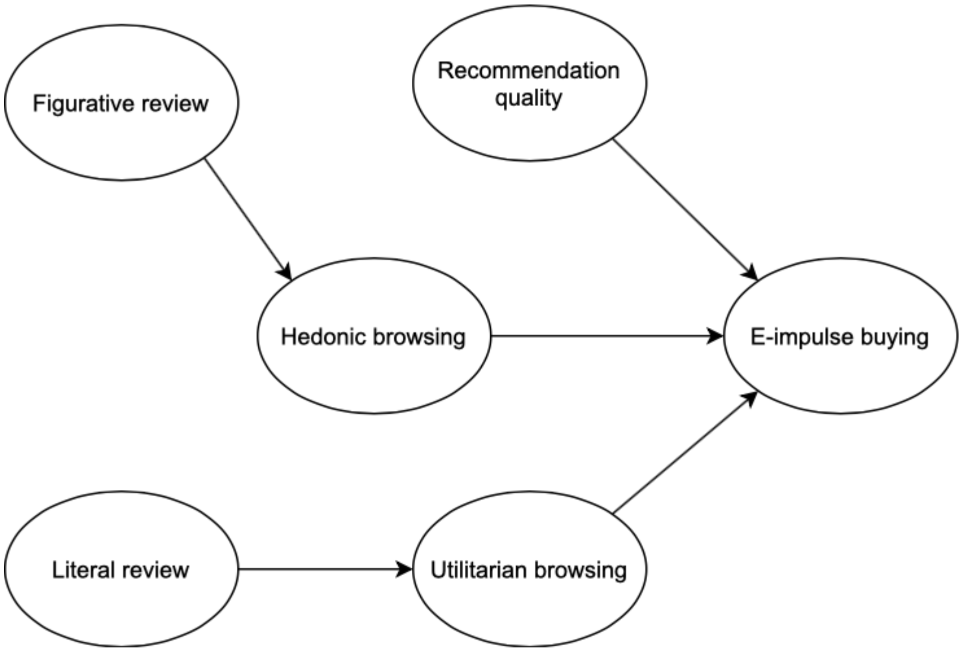
METHODS

Measurement Development and Model

The data measurement items are adopted from extant research. Park et al. (2012) modified existing scales to measure (a) web browsing encompassing both utilitarian and hedonic aspects and (b) e-impulse buying in their online apparel shopping research. One of the existing scales modified was the personal shopping value (PSV) scale developed by Babin et al. (1994) to assess "consumers’ shopping experiences along hedonic and utilitarian value” (Picot-Coupey et al., 2021, p. 578), which has been widely applied in research. Recently, Picot-Coupey et al. (2021) corroborated the PSV scale and extended the scale to two contemporary shopping environments: online websites and mobile apps. To date, Babin et al.’s (1994) PSV scale remains among the most applied value scales in research. We have chosen to adapt Park et al.’s (2012) more recent application of scales to measure utilitarian browsing and hedonic browsing. We adapt Park et al.’s scales to measure e-impulse buying as well.

We deploy Kronrod and Danziger's (2013) figurative and literal hotel descriptions to measure consumers' perceptions of hotel descriptions in customer reviews and to understand how they persuade consumers to buy trips and travel packages on the web. To measure consumers' perceptions of recommendations offered by recommender systems, we adapt Roudposhti et al.'s (2018) scale on recommendation quality.

Figure 1: The hypothesized model for e-impulse buying on travel booking websites



Sample and Data Collection

A survey was conducted on Amazon Mechanical Turk from May 5, 2021 to May 12, 2021. The sample included individuals who vacation at different intervals: vacation every month, vacation every quarter, and vacation every year. Participants were adult residents of the United States. We received 25 responses. After data cleaning, the sample was N=20 for analysis.
The sample includes more male participants (55%) than female participants (45%). The majority of participants are married (65%). Also, the majority of participants have a bachelor's degree (65%). Dominant age groups are 30-39 (40%) and 40-49 (25%).

Data Analysis

The scope of the current study is limited to a small sample size (N=20). Structural equation modeling (SEM) was used for examining the hypothesized model and to investigate the relationships between the constructs (Figure 1). SMART PLS 3.0 software was used for the analysis. The reliability of the data collection tool was verified with Cronbach's alpha and composite reliability measures. Composite reliability was verified by comparing the square root of the average variance extracted (AVE) with inter-construct correlations. Convergent validity was verified with the AVE. We used the standard cutoff values.

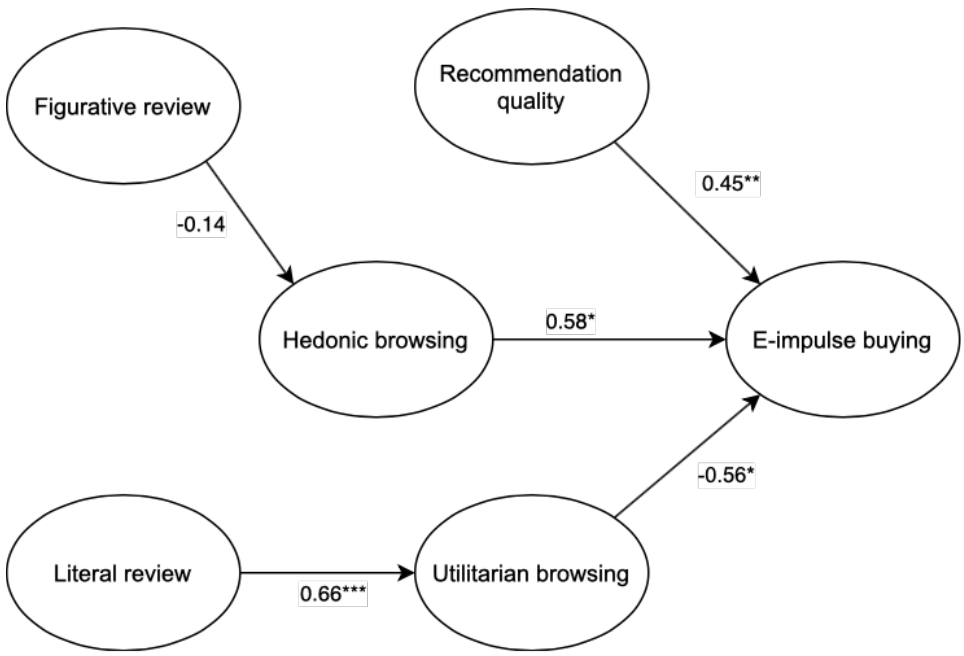
RESULTS

Path analysis provides evidence in favor of the hypothesized relationships (Table 1). All independent constructs have significant coefficients and provide support for the hypothesized relationship with the dependent construct (e-impulse). Recommendation quality (β = 0.45, p < 0.01) and hedonic browsing (β = 0.58, p <.05) have highly significant positive effects on e-impulse buying whereas utilitarian browsing (β = -0.56, p < 0.05) has a significant negative effect on e-impulse buying. Figurative customer review (β = -0.14) meets the hypothesized relationship requirement; however, the standardized coefficient is nonsignificant. Literal customer review (β = 0.66, p < 0.001) has a highly significant positive effect on utilitarian browsing. Figure 2 shows the path relationship and standardized estimates between dependent and independent variables/constructs.

Table 1: PLS-SEM coefficients (standardized)

Standardized	Coef.	Std. Dev.	t	p>t	[95% Conf. Interval]	
E-impulse						
Figurative review	-0.14	0.29	0.52	0.577	-0.54	0.65
Literal review	0.66	0.15	4.40	0.000	0.36	0.91
Recommendation quality	0.45	0.16	2.73	0.009	0.08	0.71
Hedonic browsing behavior	0.58	0.21	2.77	0.013	0.14	0.91
Utilitarian browsing behavior	-0.56	0.21	2.66	0.012	-0.10	-0.13

Figure 2: PLS-SEM path and coefficients (standardized) $*p$<0.05; $**p$<0.01; $***p$<0.001



DISCUSSION AND CONCLUSION

Our research examines the effect of supporting services in e-commerce, which has received little attention in extant literature. Only one hypothesized effect is nonsignificant. Figurative review did not encourage hedonic browsing. Also, recommendation novelty, diversity, and accuracy are shown to have a significant effect on recommendation quality in Roudposhti et al.'s (2018) study. Due to limitations a sample size presents, we tested only the effect of recommendation quality on e-impulse buying. In extended research, we will explore the dimensions of novelty, diversity, and accuracy in product recommendations. A larger sample will allow testing more links in the model with greater power.

REFERENCES

Babin, B.J., Darden, W.R., & Griffin, M. (1994). Work and/or fun: Measuring hedonic and utilitarian shopping value. *Journal of Consumer Research, 20*(4), 644–656.

Cenfetelli, R.T., Benbasat, I., & Al-Natour, S. (2008). Addressing the what and how of online services: Positioning supporting-services functionality and service quality for business-to-consumer success. *Information Systems Research, 19*(2), 161–181.

Chan, T.K.H., Cheung, C.M.K., & Lee, Z.W.Y. (2017). The state of online impulse-buying research: A literature analysis. *Information & Management, 54*(2), 204–217.

Choi, J., Lee, H.J., & Kim, Y.C. (2011). The Influence of social presence on customer intention to reuse online recommender systems: The roles of personalization and product type. *International Journal of Electronic Commerce, 16*(1), 129–153.

Elwalda, A., & Lü, K. (2016). The impact of online customer reviews (OCRs) on customers' purchase decision: An exploration of the main dimensions of OCRs. *Journal of Customer Behaviour, 15*(2), 123–152.

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545–560.

Hazari, S., Bergiel, B.J., & Sethna, B.N. (2017). Hedonic and utilitarian use of user-generated content on online shopping websites. *Journal of Marketing Communications, 23*(6), 572–591.

Hong, H., Kim, H., & Lennon, S.J. (2018). The effects of perceived quality and usefulness of consumer reviews on review reading and purchase intention. *Journal of Consumer Satisfaction, Dissatisfaction & Complaining Behavior, 31*, 97–115.

Hostler, R.E., Yoon, V.Y., Guo, Z., Guimaraes, T., & Forgionne, G. (2011). Assessing the impact of recommender agents on on-line consumer unplanned purchase behavior. *Information & Management, 48(*8), 336–343.

Jiménez, F.R., & Mendoza, N.A. (2013). Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products. *Journal of Interactive Marketing, 27*(3), 226–235.

Kim, S., & Eastin, M.S. (2011). Hedonic tendencies and the online consumer: An investigation of the online shopping process. *Journal of Internet Commerce*, *10*(1), 68–90.

Kronrod, A., & Danziger, S. (2013). "Wii will rock you!" The use and effect of figurative language in consumer reviews of hedonic and utilitarian consumption. *Journal of Consumer Research, 40*(4), 726–739.

Liao, C., To, P., Wong, Y., Palvia, P., & Kakhki, M.D. (2016). The impact of presentation mode and product type on online impulse buying decisions. *Journal of Electronic Commerce Research, 17*(2), 153–168.

Lo, L.Y., Lin, S., & Hsu, L. (2016). Motivation for online impulse buying: A two-factor theory perspective. *International Journal of Information Management*, *36*(5), 759–772.

Park, E.J., Kim, E.Y., Funches, V.M., & Foxx, W. (2012) Apparel product attributes, web browsing, and e-impulse buying on shopping websites. *Journal of Business Research, 65*(11), 1583–1589.

Picot-Coupey, K., Krey, N., Hur´e, E., & Ackermann, C. (2021). Still work and/or fun? Corroboration of the hedonic and utilitarian shopping value scale. *Journal of Business Research, 126*, 578–590.

Roudposhti, V.M., Nilashi, M., Mardani, A., Streimikiene, D., Samad, S., & Ibrahim, O. (2018). A new model for customer purchase intention in e- commerce recommendation agents. *Journal of International Studies, 11*(4), 237–253.

Sela, A., & Berger, J. (2012). How attribute quantity influences option choice. *Journal of Marketing Research, 49*(6), 942–953.

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems, 86*, 48–60.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly, 31*(1), 137–209.

DECISION SCIENCES INSTITUTE

The Shifting Gears of Infobesity: Does Frequency Drive Degree?

Ojaswi Malik
The University of Hong Kong
Email: ojaswim@connect.hku.hk

Prasanna Karhade
University of Hawaii Manoa
Email: karhade@hawaii.edu

Abhishek Kathuria
Indian School of Business
Email: abhishek_kathuria@isb.edu

ABSTRACT

Infobesity is the condition wherein firms and decision makers collect more information than they need or they can efficiently use. Recent research has commenced examining the antecedents of infobesity in organizations. However, we lack an understanding of the relationship between two important facets of experiencing infobesity: frequency and degree. For this purpose we leverage an emergent research design that harnesses decision tree induction to reveal tacit patterns. Our investigation uncovers interactions of antecedents of infobesity across multiple levels of enquiry. We conclude that the frequency of encountering infobesity drives the perception of the degree at which infobesity will be experienced.

KEYWORDS: Infobesity, Information Overload, Induction, Abduction, Enterprise Systems, ERP

INTRODUCTION

Organizations across a variety of industries face infobesity and struggle to identify important data in a timely manner. For many firms, infobesity has become a blind spot as they have become desensitized to being overwhelmed with information. Infobesity is the condition wherein firms and decision makers collect more information than they need or they can efficiently use. Infobesity limits decision makers' attention capacity thereby making it increasingly difficult to effectively use all available information (Eppler & Mengis, 2004; Hemp, 2009).

Infobesity in firms is caused by an abundance of information, which can differ in two important facets namely frequency and degree of experiencing infobesity. Recent research has begun to uncover the impact frequent interruptions and interactions can have on cognitive load (Gupta et al., 2013; Liang & Fu, 2017). However, there remains a need to study how varying frequencies of information from enterprise systems can lead to different degrees of infobesity in organizations. Moreover, while past studies (Karhade & Dong, 2021) have explained how firms perform in the presence of infobesity, there remains a gap in in comprehensively understanding the complex interplay of multiple factors residing across multiple levels of analyses which can address infobesity. A careful examination of the antecedents of infobesity is necessary so that decision makers can be equipped with the capabilities to transform excess information and make it value-adding. Hence, we aim to open the black box of infobesity by examining the complex interplay of multiple levels of predictors with frequency of infobesity, in their influence the firm's infobesity experience. By doing so we explain whether seasonal bursts of information overload can have a similar impact on the degree of infobesity as does a constant barrage of information in an organization.

Hence this study focuses on the following research question: *What impact does the frequency of infobesity have on the degree of infobesity in organizations?*

We leverage a research design that has gained credence in recent scholarship (Karhade & Kathuria, 2020; Karhade et al., 2015; Kathuria et al., 2020). Our methodological approach involves employing an induction-based, data driven analytics approach to uncover emergent patterns from the data. Decision tree induction serves as the basis for identifying tacit patterns in the data. This method allows us to reveal key pathways of predictors at different levels of analyses that elucidate the combination of different predictors that create infobesity in the firm. We are then able to make sense of these patterns to offer best plausible explanations that outline the interactions between the antecedents of infobesity.

This study uses a survey dataset collected from senior decision makers including executive level officers from a sample of more than two hundred U.S. firms. Data was collected on the degree and frequency of infobesity experienced by the firms, sourced from enterprise systems including Supply Chain Management (SCM), Customer Relationship Management (CRM), and Enterprise Resource Planning (ERP) systems. Furthermore, extensive data on the firms' partner network and firm-wide investment efforts was also collected.

Our findings confirm that the frequency of experiencing infobesity is an essential predictor of the degree with which organizations encounter infobesity. When firms rarely experience infobesity (thus, in bursts), its decision makers are not likely to find the firms' information sources to be more than what is needed or can be efficiently used. In contrast, when firms and their decision makers occasionally or frequently experience infobesity, they are likely to experience moderate to high degrees of infobesity. Moreover, this relationship is governed by the levels at which the firm implements enterprise systems, the average product lifecycle, and the firm's investment in IT. Hence, the rate at which firms encounter infobesity drives the perception of the intensity at which the overload will be experienced. These findings have implications for the attention-based view of the firm (Ocasio, 1997), and illuminate the forces controlling decision makers' limited attention.

LITERATURE REVIEW

Sources of infobesity

With the advent of information technology, firms' information environment is becoming increasingly complicated. Exposed to copious amounts of data through information systems, firms need to cope with the phenomenon of Infobesity. The increasing use of different types of information systems - such as enterprise systems, email systems – can lead to infobesity (Karhade & Dong, 2021; Stich et al., 2019). Firms are similar to information processing systems and an excess of information load which is greater than the firms' information processing capabilities can create infobesity. (Tarafdar et al., 2010). Past literature has extensively researched the predictors and impacts of infobesity (Eppler & Mengis, 2004; Hemp, 2009; Karhade et al., 2020). High levels of infobesity can overwhelm firms' decision makers and take away from their ability to take timely decisions (Hemp, 2009; Jackson & Farzaneh, 2012).

Moreover, it can elicit negative reactions such as frustration and dissatisfaction in workers (Hemp, 2009; Ragu-Nathan et al., 2008). This can have an overall detrimental effect on organizations' success and performance (Simon, 1996).

The firm's information environment is influenced by variables which exist at four major levels of analyses including the technology use level, partner-level, firm-level, and industry level.

A firm's endowments in technology affects its information environment both negatively and positively (Ramakrishnan, Kathuria, et al., 2018; Ramakrishnan, Khuntia, et al., 2018). Implementing several digital resources exposes the firm to an abundance of information which can create infobesity in the organization. On the other hand, past literature has established that IT-enabled capabilities related to information management have a positive impact on the organizations' performance such as improving productivity and profitability as well as reducing costs (Dong, He, et al., 2013; Dong, Karhade, et al., 2013a, 2013b; Mithas et al., 2011; Ramakrishnan et al., 2020; Thiagarajan et al., 2020). Implementation of enterprise systems such as Customer Relationship Management and Supply Chain Management systems, firms' decision makers are able to expand their knowledge sources by acquiring vital information from their value chain, and can thereby improve their innovation outcomes (Gómez et al., 2017; Kleis et al., 2012; Rai et al., 2006).

The organization's information environment changes in quantity and quality as the firm interacts with its partners. The firm's decision makers are incentivized to seek a large variety of additional information and knowledge in the environment through collaborative activities (Un et al., 2010). Firms can establish multiple partnerships with suppliers and customers in its value chain network to access a larger variety of information (Faems et al., 2005). Prior literature has identified the benefits of employing collaboration technologies within teams (Magni & Maruping, 2019; Maruping & Magni, 2015). However, by gaining new and diverse knowledge from its partners, the firm is more likely to face an excess load of information. This creates an additional source of information which can lead to infobesity.

A firm's digitally enabled activities and access to information sources is determined by its resource endowments such as its investment in R&D and IT (Andrade et al., 2020; Dong et al., 2021; Dong et al., 2015; Saldanha et al., 2021). Furthermore, the firm's market scope and merger and acquisition strategy influences the amount and diversity of information a firm is exposed to (Andrade Rojas & Kathuria, 2014). These firm-level predictors can have an impact on the degree and frequency with which the organization will experience infobesity.

Lastly, the culmination of advances in information technology has led to wide- spread automation and digitalization at the industry-level. With increased technical capital, industries' decision-making speed has increased, along with quicker production of high-quality products and services (Sima et al., 2020). This has led to industries becoming fast-paced and dynamic witnessing a fast pace of innovation or high clockspeed (Mendelson and Pillai 1999).

Two Facets of Infobesity: Frequency and Degree

Past research has uncovered several characteristics of information, such as the frequency, complexity, intensity and quantity, that can determine the infobesity experienced by organizations (Eppler & Mengis, 2004; Roetzel, 2019). The frequency of infobesity deals with when and how often organizations have to deal with excess information. On the other hand, the degree of infobesity is about how much information creates infobesity in the firm. The degree of infobesity measures the different levels by which information can become a source of infobesity. Prior literature has begun to study the phenomenon of frequency of infobesity such as how often social media platforms or online advertising create interruptions for users (Liang & Fu, 2017; McCoy et al., 2007). Furthermore, studies about the impact of frequency of interruptions in organizations on workers' performances have revealed the consequences of infobesity on decision makers' productivity and performance (Gupta et al., 2013).

However, there remains a gap in examining whether the two major facets of infobesity, frequency and intensity, are interdependent. By explaining the behavior of these antecedents of infobesity, firms' decision makers can be equipped with the capabilities to transform the excess information and make it value-adding. Hence, this study explores how different frequencies of information from enterprise systems creates different degrees of infobesity. We aim to open the black box of overload by comprehensively explaining the complex interplay of multiple levels of predictors, specifically frequency of infobesity, that influence the firm's infobesity experience.

METHODS

Data

This study uses survey data collected from a sample of more than two hundred U.S. firms (Karhade and Dong 2021). The survey respondents were senior executives of a sample drawn from a mix of eight high and low clockspeed (Mendelson & Pillai, 1999) industries — business services, electronics & telecommunications, transport & logistics, computer hardware & services, , food & beverages, chemicals & pharmaceuticals, retail, and energy & mining.

Measures

Dependent Variable – Degree of Infobesity

Our research design aims to build theory through the sensemaking of patterns revealed through the induction of decision trees. For this purpose, the *Degree of Infobesity* faced by firms is our key outcome variable. We collected data about the degree of infobesity faced by firms and its decision makers using SCM, CRM and ERP systems individually on a 7-point scale. Degree of infobesity was calculated as a combination of three characteristics. First, if the information collected via these three enterprise systems was more than needed by the organization. Second, if the information collected via these systems was more than could be efficiently used by the organization and its decision makers. Third, if the information collected via these three different systems was a source of information overload. Then, for each firm we took the sum of the values from the three information systems and classified firms as belonging to one of High, Medium, or Low degree of Infobesity.

Frequency of infobesity

Our key independent variable is the *Frequency of Infobesity* faced by firms and its decision makers by using SCM, CRM and ERP systems. This variable is measured as: by how often firms experience infobesity on a seven-point scale ranging from 1 (Never) to 7 (Always) from each of the three enterprise systems. For each firm we took the sum of the frequencies of infobesity from the three information systems and classified firms as belonging to one of High, Medium, or Low frequency of infobesity.

Technology-use attributes

We captured the technology use capabilities of the firm through three key enterprise systems namely ERP, SCM, and CRM systems. We collected information on the *ERP system use* by firms to manage their internal information. Firms that did not employ any ERP systems were classified as possessing *No ERP Use,* whereas firms that employed ERP systems were classified in the *ERP Use* category. For SCM and CRM systems, we collected data on the proportion of customers and proportion of suppliers connected to the systems respectively. Firms were classified into one of High/Medium/Low *SCM system use* and High/Medium/Low *CRM system use* based on the proportion of suppliers and customers connected to the systems respectively

Partner-level attributes

We captured the firms' supplier and customer partner network through the *Number of Suppliers* and *Number of Customers* respectively. Number of customers and Number of Suppliers were coded into high, medium, and low values based on top, middle, and lower one-third categories of values.

Firm-level attributes

We incorporated four firm-level attributes to improve our understanding of the antecedents of Infobesity. A firm was classified as having High/Medium/Low IT Investment based on the average spending on IT hardware, software, and services as a percentage of total sales. Similarly, firm's *R&D Investment* was categorized as High/Medium/Low based on average spending on research and development as a percentage of total sales. Moreover, the sample firms were classified as exhibiting High/Medium/Low *M&A activity*. Finally, firms were classified as having International, Regional or Domestic *Market Scopes*.

Industry-level attributes

We classified a firm in a High/Medium/Low *Clockspeed* industry if its average product lifecycle was less than 1 year; greater than one year, but less than 2 years; and more than 2 years respectively.

Methodology

We use the supervised machine learning methodology of decision tree induction to uncover complex underlying relationships in the data that otherwise are tacit and left unidentified. The induction algorithm utilizes the most informative attributes that influence the outcome to provide context-specific rules (Karhade & Shaw, 2007; Karhade et al., 2015). We use the open source Weka data mining tool for data partitioning, inducing trees, and pruning trees.

Before inducing the decision trees, we conduct the essential step of data partitioning. Two random, mutually exclusive training and testing subsamples of observations are drawn repeatedly from the data. The knowledge from the training partition is used to identify tacit connections and grow the decision trees. Then, the disjoint set of observations in the testing set are used to test the predictive accuracy of the decision rationale discovered in the training set. Post partitioning, the decision tree induction methodology involves two main steps of inducing the trees, followed by pruning the trees. We utilize the C4.5 algorithm to induce trees on the training partition. The induced trees are pruned using the testing partition. Decision tree induction has two key inputs namely (1) firms described by all information attributes and (2) infobesity experienced by firms. After employing the C4.5 induction algorithm, the output is a decision tree that unveils tacit relationships of attributes leading to similar final outcomes. The patterns can then be summarized as rules.

The C4.5 algorithm utilizes the concepts of information entropy and information gain ratio to reduce impurity in determining which attributes lead to terminal nodes or leaves. Hence, the tree induction methodology iteratively groups together firm-level observations that not only demonstrate similar information attributes, but also lead to the common final outcome. This hereby leads to the output decision trees retaining only the most informative attributes.

While the decision tree uncovers the patterns in the data, the role of the researcher remains to conduct sensemaking to best explain the relationships revealed. This requires the researchers' expertise and judgement to offer the most plausible explanations, and not confirmative logic, for the observations derived from induction (Douven, 2017). This allows us to test various choices, improve predictive performance and ultimately offer the best possible explanation to the evidence available.

FINDINGS FROM INDUCTION

The three heuristics of (1) high prediction accuracy (2) high parsimony (3) high reliability are relied on in order to maintain robustness and select the best representative decision tree. The most informative attribute is the attribute that appears at the top of the decision tree. Importance of attributes decreases as we move away from the top of the tree towards its leaves. Implications of decision tree induction lie in not only the attributes that appear in the tree, but also in the attributes that do not appear in the tree.

Frequency of Infobesity was the top-most and informative attribute, with SCM System Use being the second most important attribute for infobesity. Attributes that do not appear in the tree or appear at low levels are the predictors at the firm-level, partner-level and industry-level. These patterns lend credence to our core premise that the frequency of infobesity has substantive influence on the degree of infobesity experienced by firms and its decision makers.

It is pertinent to note that the induced trees are not exact rules used by firms' decision makers, but instead are robust approximations of the tacit underlying decision rationale. We present two main rules revealed from the induction of trees next.

Rule 1: Low frequency does not drive High Degree

Firms that experience a low frequency of infobesity are more likely to encounter low or moderate degrees of infobesity. This varies depending on the level with which firms deploy enterprise systems. We discovered that firms that implement SCM systems at a low level and face low frequencies of infobesity experience a low degree of infobesity. Furthermore, firms that deploy SCM systems at a higher level and experience low frequencies of infobesity are more likely to encounter only a moderate degree of infobesity. These findings indicate that firms that never or very rarely experience infobesity are not likely to experience high degrees of infobesity.

Rule 2: Medium and High Frequency drives High Degrees of infobesity

When a firm occasionally or frequently experiences infobesity, it is likely to experience moderate to high degrees of infobesity. This conclusion is supported by two patterns revealed in the tree.

First, we discovered that firms that face a medium frequency of infobesity tend to encounter medium or high degrees of infobesity. When facing moderate frequencies of infobesity, a high or medium degree of infobesity is determined by factors at the technology-use, industry and firm level. If the firm occasionally experiencing moderate frequency connects a high proportion of its suppliers through SCM systems, it is exposed to a high degree of infobesity. On the other hand, if the firm connects only a few suppliers by using SCM systems, it is likely to experience only a medium degree of infobesity. The tree revealed industry clockspeed to be an important attribute for determining firms' decision makers' degree of experiencing infobesity when facing a moderate frequency of infobesity and medium SCM system use. In particular, belonging to an industry with long product lifecycles, exposes firms to moderate degrees of information overload. On the other hand, firms in industries with shorter lifecycles can experience high degrees of infobesity when the firm has a high IT spending.

Second, the decision tree revealed firms that very frequently experienced infobesity also experienced high degrees of infobesity as a consequence. This implies that when facing infobesity very often, firms' sources of information cannot be of low or moderate degrees of infobesity. These findings lead to the conclusion that frequency of infobesity is an important determinant of degrees of infobesity.

DISCUSSION

The induction of decision trees allows us to identify the relationship among the combination of different predictors causing infobesity and derive context-specific rules (Dasgupta et al., 2021; Karhade et al., 2021; Kathuria et al., 2020). After inducing these two rules from the decision trees, we make sense of the patterns revealed. Doing so will allow us to identify different explanations, and ultimately arrive on the most plausible explanation for the induced findings.

The two induced rules confirm that the frequency of experiencing infobesity is an essential predictor of the degree with which organizations encounter infobesity. Our findings indicate a linear interrelationship between frequency and degree of infobesity. Rule 1 reveals that when firms rarely experience infobesity, its decision makers are not likely to find the firms' information sources to be more than what is needed or can be efficiently used. In contrast, Rule 2 highlights that when firms and its decision makers occasionally or frequently experiences infobesity, they are likely to experience moderate to high degrees of infobesity. Moreover, this relationship is governed by the levels at which the firm implements enterprise systems, the average product lifecycle, and the firm's investment in IT. Hence, the rate at which firms encounter infobesity drives the perception of the intensity at which the overload will be experienced.

This study opens the black box of infobesity by examining the relationship between its two key facets: frequency and degree. While past studies have identified several predictors of infobesity, this study delved into explaining the nuanced relationship between degree and frequency of infobesity that determine the firm's infobesity experience. This research moves the scholarship on decision sciences forward by unravelling the theoretical construct of infobesity. As the firm's information environment grows to be more complex, infobesity further cloud the already scarce resource that is firms' decision makers' attention (Ocasio, 1997). Hence, investigating the characteristics that influence the firms infobesity is a timely and pervasive matter.

This study offers plausible explanations that underlie the relationship between frequency and degree of infobesity experienced by firms. However, we recognize that while studying the sources of infobesity is necessary, it is insufficient to establish the belle cause for examining this phenomenon. In related research, we have examined the impact of infobesity on the firms' innovation outcomes. This enables us to establish the adverse consequences of infobesity on firm performance. Findings from this related research establish infobesity as an important predictor of innovation. A key insight from this work is that digital collaborations with suppliers and customers can help decision makers make effective use of abundant knowledge by filtering information about market needs to develop new products and services with enhanced features (Karhade & Dong, 2021; Khuntia et al., 2014). These findings, though beyond the scope of this manuscript, indirectly examine the impact of frequency and degree of infobesity on the firm's innovation, and thus present avenues for further examinations.

CONCLUSION

This study is the first step towards establishing a deeper shared understanding of infobesity. We recognize the growing importance of digitized social networks such as Instagram and Twitter, and the overwhelming amount of information they generate for organizations which can fundamentally alter the frequency with which decision makers experience infobesity (Andrade Rojas, Kathuria, et al., 2021). We believe this paves the way for acquiring deeper insights to study infobesity in organizations. While our data was collected from one decision maker from each firm, we believe future research can focus on understanding the application of infobesity at different levels, such as team-levels, within the organization. Moreover, crises such as the pandemic have exposed organizations to high frequencies of information in a relatively short period of time (Vijaykar & Karhade, 2021; Vijaykar et al., 2021). This creates opportunities for further examination of the impact of frequency on degree of infobesity during crises at a global level (Andrade Rojas, Saldanha, et al., 2021; Khuntia et al., 2021).

REFERENCES

Andrade, M., Saldanha, T., Khuntia, J., Kathuria, A., & Boh, W. (2020). Overcoming Deficiencies for Innovation in SMEs: IT for Closed Innovation versus IT for Open Innovation. *Proceedings of the International Conference on Information Systems (ICIS) 2020*.

Andrade Rojas, M. G., & Kathuria, A. (2014). Competitive Brokerage: External Resource Endowment and Information Technology as Antecedents. *Academy of Management Proceedings 2014*(1), 15643.

Andrade Rojas, M. G., Kathuria, A., & Konsynski, B. (2021). Competitive Brokerage: How Information Management Capability and Collaboration Networks Act as Substitutes. *Journal of Management Information Systems, Forthcoming*.

Andrade Rojas, M. G., Saldanha, T., Khuntia, J., Kathuria, A., & Boh, W. (2021). Overcoming Innovation Deficiencies in Mexico: Use of Open Innovation through IT and Closed Innovation through IT by Small and Medium Enterprises, . *Hawaii International Conference on System Sciences, 2021*, 617.

Dasgupta, A., Karhade, P., Kathuria, A., & Konsynski, B. (2021). Holding Space for Voices that Do Not Speak: Design Reform of Rating Systems for Platforms in GREAT Economies. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2564.

Dong, J., Karhade, P., Rai, A., & Xu, X. (2021). How firms make information technology investment decisions: Toward a behavioral agency theory. *Journal of Management Information Systems, 38*(1), 29-58.

Dong, J. Q., He, J., & Karhade, P. (2013). The Penrose effect in resource investment for innovation: Evidence from information technology and human capital. *European Conference on Information Systems Proceedings*.

Dong, J. Q., Karhade, P., Rai, A., & Xu, S. X. (2013a). Dynamic adjustment of information technology, corporate governance, and firm profitability. *European Conference on Information Systems Proceedings*.

Dong, J. Q., Karhade, P., Rai, A., & Xu, S. X. (2013b). Information technology in innovation activity of the firm: Theory and synthesis. *European Conference on Information Systems Proceedings*.

Dong, J. Q., Karhade, P., Rai, A., & Xu, S. X. (2015). Information Technology and Innovation Outputs: The Missing Link of Search Evolution. Academy of Management Proceedings,

Douven, I. (2017). *Abduction* (Summer 2017 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/abduction/

Eppler, M. J., & Mengis, J. (2004). The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. *Information Society, 20*(5), 325-344.

Faems, D., Van Looy, B., & Debackere, K. (2005). Interorganizational collaboration and innovation: Toward a portfolio approach. *Journal of Product Innovation Management, 22*(3), 238-250.

Gómez, J., Salazar, I., & Vargas, P. (2017). Does information technology improve open innovation performance? An Examination of Manufacturers in Spain. *Information Systems Research, 28*(3), 661-675.

Gupta, A., Li, H., & Sharda, R. (2013). Should I send this message? Understanding the impact of interruptions, social hierarchy and perceived task complexity on user performance and perceived workload. *Decision Support Systems, 55*(1), 135-145.

Hemp, P. (2009). Death by information overload. *Harvard business review, 87*(9), 82-89, 121.

Jackson, T. W., & Farzaneh, P. (2012). Theory-based model of factors affecting information overload. *International Journal of Information Management, 32*(6), 523-532.

Karhade, P., & Dong, J. Q. (2021). Innovation outcomes of digitally enabled collaborative problemistic search capability. *MIS quarterly, 45*(2), 693-718.

Karhade, P., & Kathuria, A. (2020). Missing Impact of Ratings on Platform Participation in India: A Call for Research in GREAT Domains. *Communications of the Association for Information Systems, 47*(1), 19.

Karhade, P., Kathuria, A., & Konsynski, B. (2021). When Choice Matters: Assortment and Participation for Performance on Digital Platforms. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 1799.

Karhade, P., Kathuria, A., Malik, O., & Konsynski, B. (2020). Digital Platforms and Infobesity: A Research Agenda. Workshop on E-Business,

Karhade, P., & Shaw, M. (2007). Rejection and selection decisions in the IT portfolio composition process: An enterprise risk management based perspective. *AMCIS 2007 Proceedings*, 221.

Karhade, P., Shaw, M. J., & Subramanyam, R. (2015). Patterns in information systems portfolio prioritization. *MIS Quarterly, 39*(2), 413-434.

Kathuria, A., Karhade, P. P., & Konsynski, B. R. (2020). In the realm of hungry ghosts: Multi-level theory for supplier participation on digital platforms. *Journal of Management Information Systems, 37*(2), 396-430.

Khuntia, J., Kathuria, A., Andrade Rojas, M. G., Saldanha, T., & Celly, N. (2021). How Foreign and Domestic Firms Differ in Leveraging IT-enabled Supply Chain Information Integration in BOP Markets: The Role of Supplier and Client Business Collaboration. *Journal of Association of Information Systems, 22*(3).

Khuntia, J., Saldanha, T., & Kathuria, A. (2014). Dancing in the tigers' den: MNCs versus local firms leveraging IT-enabled strategic flexibility. *Proceedings of the International Conference on Information Systems (ICIS)*

Kleis, L., Chwelos, P., Ramirez, R. V., & Cockburn, I. (2012). Information technology and intangible output: The impact of IT investment on innovation productivity. *Information Systems Research, 23*(1), 42-59.

Liang, H., & Fu, K.-w. (2017). Information overload, similarity, and redundancy: Unsubscribing information sources on Twitter. *Journal of Computer-Mediated Communication, 22*(1), 1-17.

Magni, M., & Maruping, L. (2019). Unleashing innovation with collaboration platforms. *MIT Sloan Management Review, 60*(3), 1-5.

Maruping, L. M., & Magni, M. (2015). Motivating employees to explore collaboration technology in team contexts. *MIS Quarterly, 39*(1), 1-16.

McCoy, S., Everard, A., Polak, P., & Galletta, D. F. (2007). The effects of online advertising. *Communications of the ACM, 50*(3), 84-88.

Mendelson, H., & Pillai, R. R. (1999). Industry clockspeed: Measurement and operational implications. *Manufacturing & service operations management, 1*(1), 1-20.

Mithas, S., Ramasubbu, N., & Sambamurthy, V. (2011, 03/01). How Information Management Capability Influences Firm Performance. *MIS Quarterly, 35*, 237-256. https://doi.org/10.2307/23043496

Ocasio, W. (1997). Towards an attention - based view of the firm. *Strategic management journal, 18*(S1), 187-206.

Ragu-Nathan, T., Tarafdar, M., Ragu-Nathan, B. S., & Tu, Q. (2008). The consequences of technostress for end users in organizations: Conceptual development and empirical validation. *Information Systems Research, 19*(4), 417-433.

Rai, A., Patnayakuni, R., & Seth, N. (2006). Firm performance impacts of digitally enabled supply chain integration capabilities. *MIS Quarterly*, 225-246.

Ramakrishnan, T., Kathuria, A., & Khuntia, J. (2018). Business Analytics Capability and Supply Chain Management. *Proceedings of the Americas Conference on Information Systems (AMCIS)*.

Ramakrishnan, T., Kathuria, A., & Saldanha, T. J. (2020). Business Intelligence and Analytics (BI&A) Capabilities in Healthcare. In *Theory and Practice of Business Intelligence in Healthcare* (pp. 1-17). IGI Global.

Ramakrishnan, T., Khuntia, J., Kathuria, A., & Saldanha, T. J. (2018). Business intelligence capabilities. In *Analytics and Data Science* (pp. 15-27). Springer.

Roetzel, P. G. (2019). Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. *Business research, 12*(2), 479-522.

Saldanha, T., Kathuria, A., Khuntia, J., & Konsynski, B. (2021). Ghosts in the Machine: How Marketing and Human Capital Investments Enhance Customer Growth when Innovative Services Leverage Self-Service Technologies. *Information Systems Research, (in press)*.

Sima, V., Gheorghe, I. G., Subić, J., & Nancu, D. (2020). Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review. *Sustainability, 12*(10), 4035.

Simon, H. A. (1996). Designing organizations for an information-rich world. *International Library of Critical Writings in Economics, 70*, 187-202.

Stich, J.-F., Tarafdar, M., Stacey, P., & Cooper, S. C. (2019). Appraisal of email use as a source of workplace stress: A person-environment fit approach. *Journal of the Association for Information Systems, 20*(2), 2.

Tarafdar, M., Tu, Q., & Ragu-Nathan, T. (2010). Impact of technostress on end-user satisfaction and performance. *Journal of Management Information Systems, 27*(3), 303-334.

Thiagarajan, R., Khuntia, J., Kathuria, A., & Saldanha, T. J. (2020). An Integrated Model of Business Intelligence & Analytics Capabilities and Organizational Performance. *Communications of the Association for Information Systems, 46*(1), 31.

Un, C. A., Cuervo‐Cazurra, A., & Asakawa, K. (2010). R&D collaborations and product innovation. *Journal of Product Innovation Management, 27*(5), 673-689.

Vijaykar, S., & Karhade, P. (2021). Remote Virtual Workplaces in the Pandemic: The Case of IT-enabled Service Leadership. *Pacific Asia Conference on Information Systems, 2021*.

Vijaykar, S., Karhade, P., & Gupta, M. (2021). Work-From-Home vs. Work-At-Home: A Strategic Conundrum in the Digital Age. *Americas Conference on Information Systems, 2021*.

DECISION SCIENCES INSTITUTE

Supply Chain Strategies for Lithium-Ion Battery Production

Nick Heisserer
Central Lakes College
Email: nick.heisserer@clcmn.edu

ABSTRACT

The purpose of this research is to determine which operational strategies to lower the supply chain costs of battery production for electric vehicles (EV) have the most support from academic research. The article introduces this topic with five case studies that suggest that the largest impediment to customer adoption of EV's are due to higher prices than comparable internal combustion engine (ICE) vehicles due to high battery production costs (Rossini, 2016; Klug, 2013). A review of research literature was conducted to determine what opportunities to reduce the costs of battery production in the supply chain exist. Research literature indicated that five methods were most effective at reducing supply chain costs: (1) Recycling & remanufacturing battery materials; (2) Increasing the diversity of the battery component supply chain; (3) Changing the design of battery components to use more abundant materials; (4) Improving supply chain transparency to better monitor demand forecasting; (5) Decreasing supply chain distance. To determine which methods were supported most frequently by research literature, a quantitative analysis was conducted. The analysis determined that the methods supported most frequently by the research included changing the design of battery components to use more abundant materials, improve supply chain transparency to better monitor demand forecasting, and recycling/remanufacturing of battery components. Findings from this research can help to inform the decisions of battery producers and EV startups to best utilize their resources to lower the costs of battery production and make EVs less expensive to produce.

KEYWORDS: Lithium-Ion batteries, Supply Chain, Strategies, Production, Electric Vehicles

INTRODUCTION

The purpose of this research is to determine which operational strategies to lower the supply chain costs of battery production for electric vehicles (EVs) have the most support from academic research. Many economists, scientists, and environmentalists believe that the transportation sector is on the brink of a significant shift to electric vehicles (EVs) (Saxena, 2015). EVs hold the promise to provide better performance, improved reliability, and require less maintenance than traditional internal combustion engine (ICE) vehicles while not emitting any pollution. New EV startups like Taiga Motors, Dyson Inc, Tesla Inc, Lordstown Motors, and Lucid have introduced new vehicles with hopes to disrupt existing market incumbents which sell internal combustion engine (ICE) vehicles that are not transitioning to EVs as quickly. However, because these EV startups are new market entrants, they have weaker supply chain capabilities. Supply chains of EV startups often lack information, resources, technical components, and manufacturing processes in their supply chains to compete with the market incumbents (MacDuffie, 2018).

One of the most critical factors in the EV supply chain is electric batteries. Batteries are the most expensive component of EVs and the reason that EVs are still more expensive than traditional internal combustion vehicles (Rossini, 2016; Klug, 2013). Lithium-ion batteries are the preferred battery to be used in EVs due to their low weight, high power capacity, and rechargeability. However, these batteries are expensive to produce. Numerous high costs in the supply chain, such as research and development, procurement of resources, the shipment of battery materials, and design of the batteries are challenges that must be solved to lower the costs of EVs (Betancourt-Torcat et al., 2019). The price of lithium-ion batteries has fallen 73% since 2013 (Gibson, 2019) due to economies of scale and less expensive battery materials. Currently, the price per kilowatt-hour battery in 2020 is $137/kWh (Henze, 2020). Most experts believe that the price of batteries must decline to below $100kWh for EVs to be price competitive with ICE vehicles (Henze, 2020). For battery costs to decline below $100 kWh, supply chain costs must decline, and efficiencies must improve. To illustrate the successes and challenges in supply chain operations, case studies of Taiga Motors, Dyson Inc., Tesla Inc., Lordstown Motors, and Lucid Inc will be provided to better understand the current state of affairs for EV startups.

Case Study 1: Taiga Motors

Taiga motors is an electrical transportation startup that intends to sell battery-electric snowmobiles and personal watercraft to the public (About Us., 2021). The mission of the company is “Sustainable access to the great outdoors through no-compromise engineering”. The company began in 2015 with a few engineers who wanted to create EVs that offered experiences that performed better than gasoline vehicles and were better for the environment. Taiga controls its sales and distribution channels and plans to sell its product directly to consumers.

In 2018 Taiga released its first snowmobile prototype the TS2. The snowmobile could accelerate from 0-60 mph in 3 seconds and travel 62 miles of range on a single charge (Lambert, 2018). One year later, Taiga launched three models available for preorder for the consumer with starting prices of $15,000 (Lambert, 2019). Next, Taiga issued a press release that they were working on a new all-electric personal watercraft using some of the same battery components in the snowmobile platform (Taiga Team, 2020). This watercraft is named the Taiga Orca. Taiga hopes to sell 400 Orcas in 2022 for $25,000 each. They cannot produce any more Orcas due to a lack of battery components in their supply chain (Bubbers, 2020). For its snowmobile business, Taiga had originally announced that it would ship its snowmobiles to customers during the winter of 2020. Due to supply chain challenges, it now hopes to begin shipments in the summer of 2021 (Bubbers, 2020). Taiga Motors exemplifies the challenges that all EV manufacturers face due to immature production processes (Kalaitzi, M. 2019).

Case Study 2: Dyson Inc

Dyson Inc was an established vacuum manufacturer with hopes to enter a new market with a new product. Unlike Taiga motors, Dyson was an established brand with a solid financial position in the mature vacuum cleaner market. On September 26th, 2017, in a memo to Dyson employees, James Dyson, CEO formally announced that Dyson was going to attempt to enter the EV market. Dyson planned to leverage its existing manufacturing and design capabilities that were used to create electric motors in vacuum cleaners and other components and utilize them to make EVs (Dyson, 2016). In 2017, Dyson had assigned 400 employees to the products and hoped to launch a vehicle by 2020. Dyson also announced that it had formed a supply chain partnership with battery producer Sakti3, which was developing solid-state batteries which were reported to have twice the energy density of current lithium-ion vehicles and could be charged much more quickly (Dyson, 2016). These solid-state batteries also would be safer to operate and less expensive to produce. Next, Dyson confirmed it was investing 2.7 billion USD to bring 3 electric vehicles to market by 2020 but with traditional lithium-ion batteries as its solid-state battery production was not yet ready (Campbell & Pooler, 2019). The company planned to assemble the vehicle near its relocated corporate offices in Singapore. Dyson was originally a U.K.-based company that chose to relocate to Singapore due to its proximity to its supply chain networks and lower labor costs. Unfortunately, all these plans failed to materialize. In October 2019, the company announced that it had ceased its electric car operations due to high production costs which would make it unable to create a large volume of electric vehicles able to be sold at affordable prices.

Investigations into Dyson's failure have concluded that its largest initial investment in Sakti3 was not able to produce solid-state batteries at scale at lower prices, and its supply chain was underdeveloped and not efficient (Levine, 2017). Dyson was unable to develop a supply chain that was able to provide the economies of the scale, production efficiencies, plentiful resources, manufacturing capability, and a design necessary to produce an electric vehicle under USD 150,000. They had hoped to sell their EVs at a target price of USD 50,000 per vehicle.

Case Study 3: Tesla Motors

Compared to other EV startups, Tesla has a much higher level of market capitalization, innovative technologies, and economies of scale (Stringham et al., 2015). Tesla Inc. currently has three product divisions: EVs, solar panels, and battery energy storage devices. These three divisions all allow Tesla to access large amounts of capital, build an extensive supply chain network, and use economies of scale to reduce costs. In 2020, Tesla was the largest selling automotive business of battery elective vehicles (BEVs) globally. Research articles have attributed a variety of factors for Tesla's success including producing new and superior products utilizing the attacker's advantage and cultivating a passionate and loyal customer base (Thomas & Maine, 2019).

Tesla was founded in 2003 by Martin Eberhard and Marc Tarpenning. Elon Musk was an early investor of Tesla who originally joined the company to help design the Tesla Roadster. In Tesla's beginnings, it chose a unique strategy that was different from previous failed EV ventures. Tesla's approach towards its EV was more like a luxury consumer electronics company than an automobile company. Tesla's leaders believed that the best way to enter the automobile market would be to sell a premium product first at the high end of the market. Creating a premium product would create an aspirational product that consumers would highly desire. Because Tesla did not have an advanced supply chain, it could not produce a large number of vehicles. Pricing the vehicle high would allow Tesla to generate high margins from low volume sales and invest revenues back into its supply chain. As Tesla's supply chain became more robust and efficient, it could have the capabilities to produce a slightly less expensive product to be sold at higher volumes. This process would repeat until Tesla could sell a low margin, high volume product to a large number of consumers (Thomas & Maine, 2019).

Pursuing the plan to create their first EV as a luxury item maximized Tesla's strengths of design while minimizing its weakness of an underdeveloped supply chain which would have made it difficult to compete on price alone. Tesla's first vehicle produced was the Roadster priced at

USD 110,000. Tesla utilized the Roadster to generate interest and raise the capital to begin working on more affordable luxury products to sell. Two luxury products were created – the “Model S” and “Model X”. These vehicles were not yet high volume mass-market products, but were priced lower than the Roadster to appeal to more consumers. Utilizing the revenues from these products, Tesla continued to reinvest in its supply chain and production processes to further reduce costs. As a result, it was able to manufacture and sell the Model 3, a mainstream electric car for approximately USD 40,000 (Stringham et al., 2015). Tesla is now diversifying its product portfolio by selling the Model Y, a small SUV, and soon to be releasing the CyberTruck, a larger pickup truck (Thomas & Maine, 2019). While Tesla is still a premium brand, it has plans to sell a subcompact vehicle for USD 25,000, which would appeal to middle-class consumers in developed countries (Lampert, 2021). While still less than 1% of global auto sales worldwide, Tesla has transformed itself from a new market startup to the dominant business in the electric vehicle market. Tesla provides an example for other electric vehicle startups that it is possible to enter a new market with dominant incumbents and achieve success.

Case Study #4 Lordstown Motors

Lordstown Motors is an EV startup with a mission to create and sell an electric pickup work truck. The company was founded in 2018 by Steve Burns. The company was launched out of the Workforce Automotive Group which converted ICE pickup trucks into electric trucks. Workforce owns a 10% equity stake in the company and has contributed much of its intellectual property to Lordstown Motors. Lordstown’s first vehicle for sale will be the Lordstown Endurance, an all-wheel-drive electric pickup. Originally, the Endurance was scheduled to be released in late 2020, but this has been delayed to late fall of 2021 (Szymokowski, 2020).

Before 2020, a frequent challenge for Lordstown had been to acquire funding to continue operations. In 2020, Lordstown completed a reverse merger with DiamondPeak Holdings in October 2020, and as a result is now valued at USD 1.6 billion (O’Kane, 2020). Lordstown's business strategy is unique because its first truck will only be sold as a fleet vehicle to businesses and not directly to consumers. By taking this approach, it hopes to save on marketing costs and sell in larger quantities. Lordstown’s sales strategy is to appeal to businesses that calculate the total cost of ownership instead of the initial purchase price of a vehicle. Because EVs have lower maintenance and operation costs compared to ICE vehicles, Lordstown is attempting to promote itself as a value brand to fleet services and businesses (Mceachern, 2019).

Lordstown has priced its first vehicle to sell at USD 52,000, which is similar in price to the cost of an ICE truck. To do so, Lordstown is also utilizing some unique technologies to reduce prices. Its electric truck has a simplified motor system that utilizes four electric hub motors in each of its wheels. It is also sourcing a large number of off-the-shelf components from the GM Silverado pickup truck to improve quality and seek the economies of scale utilized by GM due to its mature supply chain (Truett, 2020). The company is also partnering with LG Chem to create a battery assembly plant to reduce costs in the supply chain of batteries (Secard, 2019). At the end of 2020, Lordstown reported 20,000 fleet preorders (Truett, 2020). Lordstown Motors’ success will depend upon its abilities to fulfill them.

Case Study #5 Lucid Motors

Lucid Motors is an EV startup based in Newark California. The company is currently developing its first electric car, the Lucid Air. Lucid was originally founded in 2007 under the name Atieva (Harris, 2016). Atieva was initially focused on building electric vehicle batteries and power chains. In 2016 the company rebranded itself to become Lucid and began planning its first vehicle, the Lucid Air. Lucid entered the EV market because it believed it was able to produce a superior lithium-ion battery and electric drive chain which would provide it with a competitive advantage over other EV startups.

Lucid began constructing a large manufacturing plant in Arizona in 2016. At the plant, it hoped to produce 380,000 cars per year (Ludlow, 2020). Like other EV startups, Lucid required further investment to continue operations (Ohnsman, 2016) and it received an investment of over USD 1 billion from the Public Investment Fund of Saudi Arabia (Evarts, 2019). Lucid expects that this funding will allow it to complete the Lucid Air as well as the first phase of construction of its manufacturing plant in Arizona.

Lucid's differentiator appears to be its battery technology, which offers more power and about 30% more range than its competitors at the same price point. It hopes to achieve 517 miles of range and a top speed of 217 miles per hour at $69,900 per vehicle (Snyder, 2020) which is cost-competitive with a Tesla Model S. The company has also released an SUV concept in 2020 as a second planned vehicle. However, only its Lucid Air vehicle is currently in production due to supply chain shortages (Snyder, 2020). Lucid has also expanded its battery supply chain by creating an arrangement with Samsung SDI to improve the manufacturing capacity of its unique battery designs.

Lucid's business strategy appears to be similar to Tesla's strategy of creating a vehicle first for the luxury electric vehicle market. The company hopes to leverage its battery and drivetrain technologies that it believes are superior to the technologies found in Teslas vehicles. It has not been determined yet if Lucid's technologies are better than those from other EV companies. Even if Lucid's technologies are superior, it still lacks the manufacturing capacity, supply chain, and name recognition of other EV startups. Lucid hopes to release the Lucid Air in 2021 (Lambert, 2020). The success of this product will likely determine the success of the company.

Problem Statement:

While each of these startups has experienced unique opportunities and difficulties, all have experienced challenges with their supply chain operations. These supply chain challenges will make it difficult to reduce the prices of EV battery components which will be required for these startups to capture a larger share of the automotive market. For consumers to purchase EVs, the prices of these vehicles must be at or below the price of ICE vehicles (Henze, 2020). Because the lithium-ion battery pack is the most expensive component of an EV, the best way to lower the costs of EVs is to lower the cost of batteries (Rossini, 2016). To lower the costs of batteries, the costs of battery cell production must be reduced in the supply chain. While numerous studies have been conducted to determine methods to lower supply chain production costs of electric vehicle batteries, EV startups do not have the time or resources to attempt every strategy. EV startups would benefit from understanding what strategies documented by research to reduce battery production costs exist and are most credible. By reducing the supply chain production costs of batteries, EV prices can be lowered, and their market share can increase.

Research Question:

Which methods documented by research to reduce battery production costs in the supply chain are most credible?

Theory & Hypothesis:

Innovation diffusion theory (IDT) explains how innovations are adopted (Rogers, 1962). IDT occurs as individuals learn about an innovation and its benefits through a variety of communication channels. At first, the numbers of those who utilize the innovation are small. However, as communication increases, the adoption of the innovation increases exponentially until the market is saturated and demand levels out. This creates an “S-curve” in the adoption of this innovation when graphed. Experts have predicted an “S-curve” of growth for EVs (Jia et al., 2020). It is hypothesized that this “S-curve” also exists for battery production in the EV supply chain. By reviewing available research of battery production strategies in the supply chain, this research seeks to increase the spread of innovation by determining which strategies would be most effective to lower the costs associated with battery production. To begin this analysis, a literature review will be conducted to determine which innovative battery production strategies are supported by peer-reviewed research.

LITERATURE REVIEW

One best practice listed in research is to optimize the battery supply chain to encompass the entire lifecycle of the electric vehicle (Betancourt-Torcat et al., 2019). This includes the vehicle charging network. Allowing for specific vehicle charging standards and installations will allow companies to fine-tune the size and cost of the battery to meet the required specifications of the end-user and save costs because they will not overbuild the product. For example, if cars can recharge overnight at specific times, it may require a smaller battery, thus reducing the cost of the vehicle. The challenge to control the entire lifecycle is very difficult as it requires a large amount of capital and customer consent. For example, if the manufacturer controls the entire operational process, they also must attempt to control power generation and distribution. This can be very difficult because it is expensive, complicated, highly regulated, and crosses a variety of political jurisdictions inside and outside of a country. The only company at present which is attempting this strategy is Tesla (2019) which sells the vehicle, a charger network, and home solar panel installations.

Wessnner (2013) described the process needed to build the US battery industry supply chain and relocate battery production closer to EV manufacturers. Before 2009, virtually all lithium-ion batteries were sourced from outside of the United States. This added risks of supply disruptions and higher costs. The American Recovery and Reinvestment Act provided grants to attempt to ameliorate this situation by incentivizing the construction of battery manufacturing plants. However, once the plants were constructed, there also was a new challenge – a lack of demand for the products as U.S. consumers were not purchasing battery EVs. This lack of demand only increased the costs of production as plants were not running at full capacity. The solution advocated by Wessner (2013) was increased funding and incentives for the U.S. battery manufacturing industry from the US government.

Remanufacturing is suggested as a possible solution to increasing the supply of lithium-ion batteries and lowering the cost of manufacturing (Li et al., 2018). The concept of remanufacturing creates a closed-loop supply chain model where the battery pack will be returned to the battery manufacturer at its end of life and reused or recycled for use in future batteries. Research has suggested that battery remanufacturing can increase profits by 9.81% for battery suppliers (Li et al., 2018) which can ultimately be passed on to consumers.

Several best practices in battery remanufacturing and recycling have been aided by technology internet of things (IoT) devices (Garrido Hidalgo et al., 2020). A best practice described in research is the creation of a reverse supply chain to manage battery remanufacturing. A reverse supply chain would operate as a traditional supply chain but in reverse – a seller would take the product back from the buyer at its end of life, and return it down through the supply chain for remanufacturing. An important technological prerequisite for a successful reverse supply chain is an IoT network that tracks devices to communicate materials and locations in real-time. Additional legislation is recommended to support this circular economy concept to require vehicle manufacturers to work with their supply chains to recycle and re-use components.

Another challenge to the supply chain of electric car batteries is the scarcity of materials. At present, lithium-ion batteries require raw materials such as cobalt, lithium, and nickel. Geman (2019) indicated that if current projections about battery manufacturing are correct, there could be a supply shortage of the materials required for batteries by the mid-2020s. There are solutions to this supply shortage. Battery manufacturers are experimenting with different battery compositions and chemistries (Kalaitzi et al., 2019) which could reduce the amount of expensive and rare materials such as cobalt that are found in existing batteries, thus reducing supply constraints and production costs. Due to the quickly changing science and demand forecasts for lithium-ion batteries, it is essential to have a diverse supply chain and a variety of technological options available to make the necessary number of batteries as inexpensively as possible.

Single sourcing of battery components is another significant challenge to an efficient supply chain. In *Managing Resource Dependencies in Electric Vehicle Supply Chains: a multi-tier case study* (Kalaitzi et al., 2019), the authors indicated that most battery manufacturers do not have a diverse supply chain. Most battery producers rely upon one or two vendors for critical components such as magnets and thermal sensors. Possible solutions are for battery manufacturers to diversify their resource pipelines or implement buffering strategies. Buffering strategies include a producer holding large volumes of resources needed for production in reserve to prevent supply shortages. However, keeping a large supply on hand can be risky and expensive. Another strategy that can offer similar results to buffering is to improve transparency in the supply chain by requiring all partners to share their production and inventory levels with each other to decrease the chance of interruptions due to a lack of supply.

Materials for lithium-ion batteries are sourced globally, which means that there are also geopolitical and production factors to consider your source of battery materials (Egbue & Long, 2012). One critical component of lithium-ion batteries is lithium, in which 90% of all global resources are currently found in Bolivia, Chile, China, the United States, and Argentina with more than 50% found in South America. While most estimates indicate that lithium is an abundant material, there are shortage concerns related to the ability to refine lithium and produce it as a component in batteries. Current production levels are only able to meet demand forecasted through 2025 (Egbue & Long, 2012). Also, because few countries control most of the supply and production of lithium, instability of the supply due to geopolitical issues is a concern for battery manufacturers. While lithium is not a costly component of the battery, it is a required component in all lithium-ion batteries. Stable trade agreements will provide battery producers with sufficient supply and availability to ensure their continuity of operations.

A new opportunity to create efficiencies in the battery production supply chain may be through blockchain technology. Blockchain can be defined as a "distributed data infrastructure or a method for recording data using a cryptanalytic hash function (Wang, 2019 pg. 63). Blockchain can also be used to improve inventory systems for recording assets – whether tangible, intangible, or digital. Blockchain can create redundancies for increased security against data loss. Finally, blockchain promotes transparency because no individual controls the data. One large advantage to the blockchain is that due to its transparent nature, it can improve trust between different firms in a supply chain.

Another new supply chain strategy determined effective by research is the cluster strategy. Cluster strategy utilizes the geographic proximity of low-cost supply components in the supply chain to create a competitive advantage (Ikram, 2018). An example of a cluster strategy could be for a battery manufacturer to create a manufacturing facility near where other required materials for the battery are extracted. Clustering may also allow for segments of the supply chain to integrate around aspects of the production process. The advantage of clustering is that it realizes some of the same efficiencies gained by vertical integration but is less capital intensive because acquiring other elements of the supply chain is not necessary. This can be helpful for a startup with fewer resources.

Investment in new and more efficient technologies can also improve efficiency by finding more abundant materials required for batteries. Recent studies have identified that lithium can be recovered from geothermal brine used in geothermal power plants, thus making the harvest of lithium more available in new areas and with complimentary environmentally friendly systems and making lithium mining, less expensive, faster, safer, and with less damage to the environment (Paranthaman et al., 2017). This process would also expand the locations available for lithium harvesting. Sodium is also being investigated as an alternative to lithium-ion battery production which would reduce supply shortages and costs (Nayak et al., 2018). Alternatives are also being explored for cobalt. Cobalt is one of the most expensive materials in current EV batteries and is also the most harmful to produce. At present, most cobalt is mined from the Republic of Congo (Kumar et al., 2019), and workers are subject to poor and unsafe working conditions. Currently, cobalt is in short supply. 58% of global cobalt production is already being utilized for lithium-ion battery production (Muralidharan et al., 2020). Research is underway to create metal alloys for cobalt-free batteries. A successor to cobalt appears to an aluminum nickelate cathode. Since cobalt is the most expensive aspect of the battery, reducing or eliminating its use could significantly lower prices.

Research has demonstrated that multiple opportunities exist to reduce the requirements for expensive resources to be placed in batteries. New chemistries and alloys show much promise to lower battery prices in the future. Improvements to the supply chain management process can create efficiencies and lower costs. Because of the variety of opportunities available, battery producers may be able to create a strategy that best utilizes their strengths and minimizes their weaknesses. In the next section, the key innovation strategies from the literature will be described and discussed.

DISCUSSION

Several key strategies to lower the costs of battery production have been established by the research literature. These strategies include battery materials recycling, increasing the diversity of the battery component supply chain, changing the design of battery components to use more abundant materials, improve supply chain transparency to better monitor demand forecasting, and decreasing supply chain distance. Figure 1 describes these findings.

Figure 1. Methods to lower supply chain costs in lithium-ion battery production

1. Recycling and Remanufacturing battery materials.
2. Increasing the diversity of the battery component supply chain.
3. Changing the design of battery components to use more abundant materials.
4. Improving supply chain transparency to better monitor demand forecasting.
5. Decreasing supply chain distance.

Battery materials recycling and remanufacturing would allow rare and expensive battery components to be recycled and reused at the end of a battery's lifespan. Recycling and remanufacturing would reduce mining and refinement costs and improve the materials available. Increasing the diversity of the battery component supply chain will allow battery producers to source multiple materials from a variety of vendors. This will allow for competition to lower component prices and ensure a stable supply of materials to allow battery production to operate at maximum efficiency. Research into new and more abundant battery chemistries and alloys is occurring quickly. These new materials hold promise to reduce the component costs of batteries and the costs of production. Supply shortages can also be addressed to reduce the price of production and increase efficiencies. Improving supply chain transparency can allow for better forecasting and a more efficient and orderly production process. Finally, decreasing supply chain distance can reduce transportation costs and product delays. While other strategies to lower battery production costs exist, these five strategies were most often described and appear to hold the most opportunity for implementation in the short term.

While this information can be useful, an electric vehicle startup may not have the time or resources to act upon all of these strategies. The startup may want to focus on the most effective strategy. To do so, a quantitative analysis shall be conducted upon available supply chain research to determine the most effective strategy to lower battery production costs. By understanding which strategies are most documented in peer-reviewed research, businesses can understand which strategies are most likely to be successful to lower the cost of battery production.

METHODOLOGY

To understand which strategies are documented most frequently, a quantitative analysis of existing research will be conducted. To obtain a significant and relevant sample of research, a query of peer-reviewed articles will be conducted to include research literature between the years 2000 and 2020 utilizing the EBSCO research database. The category filters for the query will include all Business and Economics Journals. The keyword search utilized will be "Lithium-Ion Batteries & Supply Chain Management". All the research articles will be cataloged and documented. The articles will be categorized based upon the five categories summarized previously and contain one "other" category for articles that are not associated with any of the strategies. The categories will be (1) Recycling & remanufacturing battery materials; (2) Increasing the diversity of the battery component supply chain; (3) Changing the design of battery components to use more abundant materials; (4) Improving supply chain transparency to better monitor demand forecasting; (5) Decreasing supply chain distance: (6) Other. After all the articles are categorized, the data will be compiled to determine the recurrence of each category in peer-reviewed research.

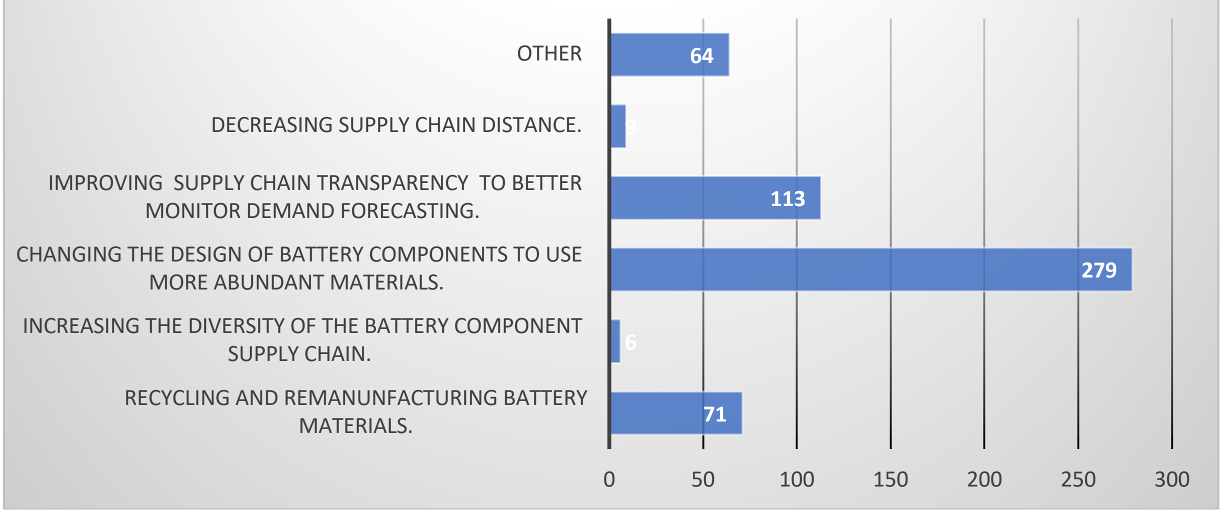
RESULTS

741 articles were retrieved using the keywords and of these articles, 199 did not apply to the battery industry. As a result, they were deleted from the report. 542 were categorized. While all six categories listed in the methodology section appeared in the research, the most common category of research was *Changing the design of battery components to use more abundant materials.* The full results are described below in Figure 2 and a graph is provided in Figure 3. The entire spreadsheet with sources is provided in Appendix A.

Figure 2: Classifications of Research: Battery Supply Chain Management

Source	**Count**	**Percent of Total**
Recycling & remanufacturing battery materials.	71	13%
Increasing the diversity of the battery component supply chain.	6	1%
Changing the design of battery components to use more abundant materials.	279	51%
Improving supply chain transparency to better monitor demand forecasting.	113	21%
Decreasing supply chain distance.	9	2%
Other	64	12%
Total	542	100%

Figure 3: Classifications of Research: Battery Supply Chain Management.



DISCUSSION & ANALYSIS

If a battery manufacturer has limited resources and desires to lower costs and/or increase efficiency in their supply chain or operations, the strategy of *Changing the design of battery components to use more abundant materials* has the most research associated with it. Research in this area often provided evidence of other battery chemistries or different metal alloys that could be used to replace more expensive materials such as cobalt and nickel. This strategy may not be appropriate for some manufacturers who do not have access to innovative battery chemistries or alloys. If the business does not have access to innovative battery production methods, two other options were also mentioned frequently in peer-reviewed research. The next largest category, *Improving supply chain transparency to better monitor demand* could also be a low-cost option by using technologies such as blockchain, or other tracking systems to share important data with other partners in the supply chain. Finally, research articles about *Recycling and remanufacturing battery materials* provided a variety of examples of companies who were able to reduce their costs and increase access to resources through a reverse supply chain production process. These three methods appeared most frequently in research and are most likely to offer the most utility to battery producers.

Research should not just be theoretical; it should be practical. Returning to Taiga Motors, Dyson Inc, Tesla Inc, Lordstown Motors, and Lucid described previously in the case study section, each could utilize their unique market positions and apply one or more of these strategies to reduce battery production costs within their supply chain. The following paragraphs will provide examples of how each startup can leverage its unique strengths and apply this research to lower battery production costs.

Taiga Motors is the only EV startup to focus on personal snowmobiles and personal watercraft. Compared to the automobile segment, the personal utility vehicle market is relatively small. Each vehicle design also requires a very specific design to protect the battery pack from snow or water. As a result, Taiga's battery design requirements are unique and may not allow it to utilize battery materials research from mass production facilities. However, because the battery packs are relatively small compared to automobiles and because they can control its sales and distribution channels, the battery recycling and remanufacturing strategy may provide it with the most utility to reduce costs of battery production. Taiga could create a battery return program to allow it to collect damaged or end-of-life battery cells and recycle or remanufacture the used batteries to create new batteries for its products and lower its costs of production. Lucid may also benefit from this strategy due to its unique design and advanced technology of its battery systems. Like Taiga, Lucid also plans to sell directly to consumers and could create a reverse supply chain. If Lucid's assertions are correct, it has a unique battery design from its competitors and could leverage a recycling and remanufacturing process of these proprietary batteries through a reverse supply chain to reduce production costs.

Dyson may have benefited from a more transparent supply chain as the main cause of its failure were due to an inefficient supply chain (Levine, 2017). By understanding its limitations and deficiencies, Dyson might have been able to attain price parity with its competitors to bring an EV to market. Lordstown Motors may also benefit from this strategy. Because it is sourcing many of its components from LG and GM including its battery components, a transparent battery supply chain may allow it to better track its costs and efficiencies to produce its truck at reasonable costs to stay within its budget and achieve a target sales price of USD 52,000.

Tesla Motors has the strongest market position and is currently the highest volume EV producer. Because of its scale, Tesla would most benefit from a strategy to reduce the costs of battery production throughout the supply chain as it already possesses the economies of scale in battery production. Tesla announced a variant of this strategy during its 2020 Annual Meeting of Stockholders and Battery Day on September 22nd, 2020, when it shared its plan to reduce battery component production costs by 50% over the next five years due to new battery designs, components, and economies of scale (Morris, 2020).

While these strategic implementations may not reflect the actual strategic positions or opportunities of each company, they demonstrate how each company can utilize this research to reduce the costs of battery production in its supply chain. They also provide an opportunity for another startup not mentioned in this study to utilize strategies of changing the design of battery components to use more abundant materials, improving supply chain transparency to better monitor demand, or recycling and remanufacturing battery materials to reduce production costs to bring their products to market. A combination of these strategies may also prove effective to reduce battery production costs in the supply chain.

Limitations

While attempts were made to include all relevant research between the years 2000 and 2020, the query did not include every research article available. Battery research articles exist in other research databases outside of EBSCO and are not always tagged in the battery or supply chain journals. Some research has not been digitized. Also, categorizing each article into five distinct categories can be problematic. Articles may contain information from multiple categories or contain information about associated topics or developments. Still, 542 articles are an adequate sample size (Maas & Hox, 2005). When an article was determined to contain multiple categories, the article was reviewed and placed in the category most aligned with its central thesis. The categorization of each article was reviewed three times to ensure accuracy by three individuals who assisted with the research. When discrepancies existed, a group discussion and review were conducted and a consensus was reached. While this research does not evaluate the quality or reliability of all the research included, the quantitative data does indicate a trend. Understanding the categories of research available and the volume of peer-reviewed research can be of use to those in the battery supply chain business to improve their operations to ultimately reduce supply chain costs.

CONCLUSION

The strategies described by research to lower battery production costs include changing the design of battery components to find less expensive materials, improving supply chain transparency, and recycling and remanufacturing batteries. While the battery supply chain and production processes are changing rapidly (Borgstedt et al., 2017), research to evaluate and verify the efficacy of these developments can be costly both in resources and time. The goal of this quantitative study was to provide all battery producers or those involved in the battery supply chain an understanding of the most credible methods available to them to reduce production costs from the large volume of research. With this information, it is hoped that these businesses will utilize the strategy that is most helpful to them to improve their supply chains, reduce battery production costs, and speed up the transition to an all-electric transportation future.

REFERENCES

About Us. (2021, January 29). Retrieved from Taiga Motors: https://taigamotors.ca/about/

Betancourt-Torcat, A., Poddar, T., & Almansoori, A. (2019). A realistic framework to a greener supply chain for EVs. *International Journal of Energy Research*, 43(6), 2369–2390. https://doi.org/10.1002/er.4373

Borgstedt, P., Neyer, B., & Schewe, G. (2017). Paving the road to electric vehicles – A patent analysis of the automotive supply industry. Journal of Cleaner Production, 167, 75–87. https://doi.org/10.1016/j.jclepro.2017.08.161

Bubbers, M. (2020, March 4). *Dashing through the snow on the Tesla of snowmobiles.* Retrieved from The Globe and Mail: https://www.theglobeandmail.com/drive/technology/article-dashing-through-the-snow-on-the-tesla-of-snowmobiles/

Ben Geman. (2019). Axios: Troubles may loom for the battery supply chain for EVs. In Axios. Newstex.

Campbell, P., & Pooler, M., (2019). Dyson cancels the electric car project. FT.com.

Dyson, J. (2016). *Memo*. Dyson Inc.

Egbue, O., & Long, S. (2012). Critical Issues in the Supply Chain of Lithium for Electric Vehicle Batteries. Engineering Management Journal, 24(3), 52–62. https://doi.org/10.1080/10429247.2012.11431947

Garrido-Hidalgo, C., Ramirez, F., Olivares, T., & Roda-Sanchez, L. (2020). The adoption of the internet of things in a circular supply chain framework for the recovery of WEEE: the case of lithium-ion electric vehicle battery packs. Waste Management (Elmsford), 103, 32–44. https://doi.org/10.1016/j.wasman.2019.09.045

Gibson, T. (2019, August 15). *Lithium-Ion Battery Prices are Declining, Powering Growth and Opportunity in the U.S. Energy Storage Market*. Retrieved from Smart Electric Power Alliance: https://sepapower.org/knowledge/lithium-ion-battery-prices-are-declining-powering-growth-and-opportunity-in-the-u-s-energy-storage-market/ 270. https://doi.org/10.1108/SCM-03-2018-0116

Gulapala Kumar, Dabbeet Kumar Prasad, & Shweta Kishore. (2019). Study of Adoption of Lithium-Ion Batteries in EVs: Scope and Challenges. Supply Chain Pulse, 10(1), 22–25.

Gonzalez, K. (2013). From the EV1 to the Chevy Volt: A re-electrification of the American automobile industry. ProQuest Dissertations Publishing.

Harris, M. (2016, October 20). *Atieva will launch its Tesla competitor by December.* Vox. https://www.vox.com/2016/10/20/13341340/atieva-launch-december

Henze, V. (2020, December 16). *Battery Pack Prices Cited Below $100/kWh for the First time in 2020, While Market Average Sits at $137/kWh*. Retrieved from BloombergNEF: https://about.bnef.com/blog/battery-pack-prices-cited-below-100-kwh-for-the-first-time-in-2020-while-market-average-sits-at-137-kwh/

Igami, M. (2017). Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998. The Journal of Political Economy, 125(3), 798–847. https://doi.org/10.1086/691524

Ikram, S. (2018). Cluster strategy and supply chain management. Benchmarking : an International Journal, 25(5), 1302–1318. https://doi.org/10.1108/BIJ-06-2015-0059

Kalaitzi, D., Matopoulos, A., & Clegg, B. (2019). Managing resource dependencies in electric vehicle supply chains: a multi-tier case study. Supply Chain Management, 24(2), 256–

Jia, J., Shi, B., Che, F., & Zhang, H. (2020). Predicting the Regional Adoption of Electric Vehicle (EV) With Comprehensive Models. IEEE Access, 8, 147275–147285. https://doi.org/10.1109/ACCESS.2020.3014851

Lambert, F. (2018, March 3). *Tesla-inspired Taiga electric snowmobile does 0-60 mph in 3 seconds*. Retrieved from electrek: https://electrek.co/2018/03/03/tesla-inspired-taiga-electric-snowmobile/

Lambert, F. (2019, June 11). *Taiga Motors launches new electric snowmobiles with impressive specs*. Retrieved from Electrek: https://electrek.co/2019/06/11/taiga-motors-electric-snowmobiles/

Lambert, F. (2020, December 1). *Lucid completes factory construction, aims to ramp up to 400,000 electric cars per year*. Electrek. https://electrek.co/2020/12/01/lucid-completes-electric-vehicle-factory-construction/

Lampert, F. (2021, February 4). *Tesla Model 3 is still the best-selling electric car in the world and it's not even close*. Electrek. https://electrek.co/2021/02/04/tesla-model-3-still-best-selling-electric-car-world-not-even-close/

LeVine, S. (2017, April 3). *Dyson has abandoned patents aimed at creating a battery-and-car juggernaut resembling Tesla*. Quartz. https://qz.com/948041/dyson-sakti3-tesla-battery-patents/

Li, L., Dababneh, F., & Zhao, J. (2018). Cost-effective supply chain for electric vehicle battery remanufacturing. Applied Energy, 226, 277–286. https://doi.org/10.1016/j.apenergy.2018.05.115

Ludlow, E. (2020, March 22). *Bloomberg*. Bloomberg - Are you a robot?. https://www.bloomberg.com/news/articles/2020-05-22/electric-cars-are-about-to-start-rolling-out-of-the-arizona-desert

Maas, C., & Hox, J. (2005). Sufficient Sample Sizes for Multilevel Modeling. Methodology, 1(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

MacDuffie, J. (2018). Response to Perkins and Murmann: Pay Attention to What Is and Isn't Unique about Tesla. Management and Organization Review, 14(3), 481–489. https://doi.org/10.1017/mor.2018.32

Mceachern, S. (2019, November 26). *Lordstown motors endurance electric pickup will start at $52,500*. GM Authority. https://gmauthority.com/blog/2019/11/lordstown-motors-endurance-electric-pickup-will-start-at-52500/

Morris, J. (n.d.). *Tesla Reveals Plan To Halve EV Battery Costs At Battery Day 2020*. Forbes. https://www.forbes.com/sites/jamesmorris/2020/09/22/tesla-reveals-plan-to-halve-ev-battery-costs-at-battery-day-2020/?sh=66eae9a57f74

Muralidharan, N., Essehli, R., Hermann, R., Amin, R., Jafta, C., Zhang, J., Liu, J., Du, Z., Meyer, H., Self, E., Nanda, J., & Belharouak, I. (2020). Lithium Iron Aluminum Nickelate, LiNixFeyAlzO2—New Sustainable Cathodes for Next-Generation Cobalt-Free Li-Ion Batteries. Advanced Materials (Weinheim), 32(34). https://doi.org/10.1002/adma.202002960

Nayak, P., Yang, L., Brehm, W., & Adelhelm, P. (2018). From Lithium-Ion to Sodium-Ion Batteries: Advantages, Challenges, and Surprises. Angewandte Chemie (International Ed.), 57(1), 102–120. https://doi.org/10.1002/anie.201703772

Ohnsman, A. (2016, November 29). *Lucid Motors Has A 1000HP Tesla Challenger; Now To Find The Cash To Build It*. Forbes. https://www.forbes.com/sites/alanohnsman/2016/11/29/lucid-motors-has-a-1000hp-tesla-challenger-now-to-find-the-cash-to-build-it/?sh=240b38183762

O'Kane, S. (2020, August 3). *GM-backed electric truck startup Lordstown motors is going public*. The Verge. https://www.theverge.com/2020/8/3/21352433/gm-lordstown-motors-electric-truck-startup-public-merger-spac

ON AN UPHILL ROLL; Tesla faces challenges in U.S. and China after ’19 ends with record sales. (2020). Automotive News, 94(6915).

Paranthaman, M., Li, L., Luo, J., Hoke, T., Ucar, H., Moyer, B., & Harrison, S. (2017). Recovery of Lithium from Geothermal Brine with Lithium–Aluminum Layered Double Hydroxide Chloride Sorbents. Environmental Science & Technology, 51(22), 13481–13486. https://doi.org/10.1021/acs.est.7b03464

Pfeffer, J., & Salancik, G. R. (1978). The external control of organizations. New York: Harper & Row.

Porter, M.E. (1997), "Competitive Strategy", *Measuring Business Excellence*, Vol. 1 No. 2, pp. 12-17. https://doi.org/10.1108/eb025476

Rogers, E. (1962). Diffusion of Innovations. New York: The Free Press

Rossini, C. (2016). A preliminary study on the changes in the Italian automotive supply chain for the introduction of EVs. Journal of Industrial Engineering and Management, 9(2), 450–486. https://doi.org/10.3926/jiem.1504

Sarkar, E., Sarkar, T., & Bharadwaj, M. (2018). Lithium-Ion Battery Supply Chain: Enabling National Electric Vehicle and Renewables Targets. Current Science (Bangalore), 114(12), 2453–. https://doi.org/10.18520/cs/v114/i12/2453-2458

Saxena, M. (2015). Charging ahead on the transition to EVs with standard 120 V wall outlets. Applied Energy, 157(C), 720–728. https://doi.org/10.1016/j.apenergy.2015.05.005

Secard, R. (2019). General Motors, LG Chem to Partner on Lordstown Battery Plant. Industry Week.

Snyder, J. (2020, August 19). *Lucid air gets over 300-kW fast charging, V2G home charging*. Autoblog. https://www.autoblog.com/2020/08/19/lucid-air-ev-charging-v2g-300kw-900v-electrify-america/

Stringham, E., Miller, J., & Clark., J.R. (2015). Overcoming Barriers to Entry in an Established Industry: Tesla Motors. California Management Review, 57(4), 85–103. https://doi.org/10.1525/cmr.2015.57.4.85

Szymkowski, S. (2020, November 16). *Lordstown motors endurance electric pickup flexes 50,000 preorders*. Roadshow. https://www.cnet.com/roadshow/news/lordstown-motors-endurance-electric-pickup-preorder-delay/

Taiga Team. (2020, July 14). *News*. Retrieved from Taiga Motors: https://taigamotors.ca/taiga%E2%80%AFintroduces%E2%80%AFnew%E2%80%AFelectric%E2%80%AFpersonal-watercraft-models-starting%E2%80%AFat%E2%80%AF15000/

Thomas, V., & Maine, E. (2019). Market entry strategies for electric vehicle start-ups in the automotive industry – Lessons from Tesla Motors. Journal of Cleaner Production, 235, 653–663. https://doi.org/10.1016/j.jclepro.2019.06.284

Truett, R. (2020) ENDURANCE TEST; Lordstown’s challenge: Starting deliveries of EV pickup in a year. (2020). Automotive News, 95(6958).

Wang, H. (2019). Understanding blockchain technology for future supply chains: a systematic literature review and research agenda. Supply Chain Management, 24(1), 62–84. https://doi.org/10.1108/SCM-03-2018-0148

Weimer, L., Braun, T., & Hemdt, A. (2019). Design of a systematic value chain for lithium-ion batteries from the raw material perspective. Resources Policy, 64, 101473–. https://doi.org/10.1016/j.resourpol.2019.101473

Wessner, C. (2013, August 12). Building the U.S. battery industry for electric drive vehicles : progress, challenges, and opportunities : summary of a symposium.

DECISION SCIENCES INSTITUTE

Post-Pandemic Business Continuity Planning for Biotechnology and Pharmaceutical Startups: Analysis and Recommendations

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

David Weidman
Boston University Metropolitan College
Email: dweidman@bu.edu

Yujue Tan
Boston University Metropolitan College
Email: cntanyj@bu.edu

Yueyi Ma
Boston University Metropolitan College
Email: yma243@bu.edu

Yuanfei Zang
Boston University Metropolitan College
Email: fibby@bu.edu

Zhenyan Yin
Boston University Metropolitan College
Email: yinzy@bu.edu

ABSTRACT

The impacts of COVID-19 on the R&D activities of biotechnology and pharmaceutical (bio/pharma) firms were studied and recommendations were made for startup bio/pharma firms. The survey participants included 320 professionals working in bio/pharma firms worldwide. The survey identified similarities by region and type of firm, but significant differences were noted when comparing bio/pharma startups to mature bio/pharma firms. Startup firms were less likely to have a preparedness plan and the plans that existed were less effective than those for other more mature firms. Recommendations were made for business continuity planning that would be useful for all startups.

KEYWORDS: COVID-19, pandemic, business continuity, biotech, pharma, sentiment analysis

INTRODUCTION

The impact of the COVID-19 pandemic was unprecedented throughout the world. Healthcare systems were overwhelmed, shortages of food and critical supplies took place, economies were severely impacted, businesses failed, and many workers lost their jobs. Although most firms had a preview to pandemic disruptions during the H5N1 outbreak of 2005-2006, and the H1N1 outbreak of 2009-2010, many firms were not prepared for the depth breadth of COVID-19 impacts. Businesses needed to contend with two critical differences between this pandemic and other disruptive events: its global impacts and its long lasting effects.

The research reported in this article includes the dissemination and analysis of two comprehensive surveys of biotechnology (biotech) and pharmaceutical (pharma) firms, and recommendations based on the surveys' results. We will refer to this grouping of firms as bio/pharma firms. Bio/pharma firms were targeted because their business was not as severely affected by the COVID-19 pandemic as other industries, as evidenced by continued steady job growth in the U.S. (Philippidis, 2020). The work focuses on business continuity (i.e., preparedness and response) planning for R&D supply chains at bio/pharma firms. Because the survey results showed important differences in the effectiveness of pandemic planning at startup firms, the recommendations concern business continuity management for startups. Although our focus is on bio/pharma, the results should be relevant for a startup or a small and medium-sized enterprise (SME) in other technological industries.

This paper is organized as follows. Background information on the operating environment and R&D activities within bio/pharma firms is presented, along with a brief introduction to business continuity management and resiliency planning. The surveys are then described, including the questions, response formats, and targeted audience. Next, survey results are shown and the differences in preparedness and effectiveness among startup firms are highlighted. The analysis and synthesis of survey results follows, using both quantitative and qualitative methods. Based on the survey results and current thinking in the business continuity professional community, we offer a set of recommendations targeted to startup firms.

BACKGROUND

Biotechnology firms study and experiment with living organisms to develop products that add value for humankind (Shackelford & Moris, 2016). Pharmaceutical firms develop, manufacture, and/or distribute medicines for human and animal use (Campbell, 2018, p. 2). Alshawi (2003) described the intellectually-oriented process flows that constitute bio/pharma business activities, including hypothesis-generation, discovery research, clinical trial development, and marketing authorization from regulatory authorities. Here, we focus on those activities that take place during development processes, which we will refer to as the R&D supply chain.

Bio/Pharma R&D Supply Chains

Like all industries, bio/pharma R&D supply chains face challenges. Those challenges cited as particularly problematic for bio/pharma include poor coordination, inaccurate demand forecasts, and human resource dependency (Privett & Gonsalvez, 2014). Internal and external barriers to effectiveness include high investment requirements, inadequate training expertise, complex rules and regulations, lack of evaluation systems, and poor end-customer awareness (Ding, 2018). Quality problems are especially troublesome and need to be dealt with quickly (Abdallah, 2013).

R&D supply chain disruptions of uncertain duration are especially problematic because they affect planning activities including demand forecasting (Socal et al, 2020). Some firms have mitigated the risks associated with these problems by creating a dual supply chain (Prakash, 2019). Other firms also create forms of redundancy or use new information technologies to mitigate risks (Handfield et al, 2020). Some bio/pharma professionals believe that the outsourcing of activities (Roehr, 2020) will never be back to the pre-pandemic "normal" and recommend moving to more self-sufficiency in their supply chains (Ayati et al, 2020). Other forms of risk mitigation include long-term agreements and government tax incentives (Roehr, 2020).

The COVID-19 pandemic has impacted bio/pharma global supply chains (Kilpatrick & Barter, 2020), including those supplying medicines and other products, causing a gap between supply and demand (Kumar, et al, 2020). Nearly 35% of bio/pharma firms report that their supply chain network has been disrupted due to the global COVID-19 pandemic (Cohen, 2020). Today's bio/pharma companies have extremely complex supply chains, in which third parties often play an important role. In addition to outsourcing the production of raw materials, many bio/pharma companies outsource key functions within their entire product life cycle, including R&D, clinical trials, manufacturing, sales, and marketing (Deloitte, 2013). As a result, third-party risks are exacerbated, including clinical trial activities (Atlas & Sobotka, 2013).

Characteristics of SMEs and Startups

Startups are a subset of firms classified as a SME. The definition of a SME varies across countries, based on characteristics such as number of employees and asset value (Lucky & Olusegun, 2012). The U.S. defines a SME as a firm with less than 500 employees (Grover & Suominen, 2014). Other governments define a SME differently, such as the European Union (Classen et al, 2014), the United Kingdom (Maurya et al, 2015), Australia (Chong, 2014), Nigeria (Alarape, 2007), Libya (Shehab, 2008), South Africa (Du Toit et al, 2009), and Malaysia (Husin & Ibrahim, 2014).

A SME is disproportionally affected by the myriad of risks that affect all businesses (Falkner & Heibl, 2015), in part due to their low equity ratios (Altman et al, 2010). Many small firms are funded internally and therefore disruptions can impact their financial well-being disproportionately (Cowling et al, 2020). A SME can also possess barriers that hinder its ability to implement sound quality and risk management principles (Murphy, 2016). Startup bio/pharma firms often have difficulty obtaining funds to mitigate risks due to their limited commercially-generated revenue (Andersson et al, 2010).

In contrast, a SME (or startup) tends to be flexible and responsive because it lacks the resources to complete based on economies of scale (Wagner & Burgstaller, 2015). Sullivan-Taylor and Branicki (2011) found that small firms are also advantaged because less bureaucracy can translate to rapid decision-making. Edvardsson et al. (2019) found that SMEs with effective outsourcing strategies perform better than SMEs without outsourcing. Finally, Dahlberg and Guay (2015) argue that implementing a business continuity program (BCP) can increase resiliency in a small firm.

Bio/Pharma Startup R&D Supply Chain Risks

The Business Continuity Institute (BCI, 2017) defines business continuity as the “The strategic and tactical capability of the organization to plan for and respond to incidents and business disruptions in order to continue business operations at an acceptable predefined level.” The BCI defines a related term, organizational resilience, as the “ability of an organization to anticipate, prepare for, and respond and adapt to incremental change in order to service and proposer.”

A BCP is typically comprised of three major components: (1) incident response and management, (2) technology recovery and business recovery, and (3) resumption. The critical success factors for BCP effectiveness include senior management access and accountability, an enterprise-wide and stakeholder scope, adequate resources, and the incorporation of emerging disciplines (Banasiewicz & Weidman, 2016, p.185). Several standards and guidelines exist for developing, implementing and maintaining a BCP. Some standards are industry-specific (e.g., Basel Accord, FERC, and FFEIC), while others are generally applicable (e.g., BS 25999, ISO 22301, and NFPA 1600).

Some authors have addressed resiliency in bio/pharma supply chains, but only a few focus on startups that are more vulnerable than mature firms (Haase & Eberl, 2019). In fact, Fochler (2016) reported a strong preference among bio/pharma startups to quickly license their discoveries. When considering a startup's resilience, it is important to consider their specific organizational structure (Haase & Eberl, 2019). However, new approaches are under development that use analytics to help startups identify projects that are particularly worthy of attention (Polishchuk et al, 2020).

Governments worldwide continuously adapt bio/pharma regulations in more holistic ways, including the privacy of patient and/or customer health information, clinical operations, and post-market safety reports (Yarza, 2019). Hence, regulatory risks are increasing for many bio/pharma firms, especially those that lack resources to respond effectively to regulatory changes (Van Norman, 2016). Regulatory modifications can impact product development in the presence of uncertain clinical trial schedules and registration delays (Maini, 2014). DiMasi et al. (2003) confirm that the complexity of regulatory structures results in higher R&D costs in startup firms.

Another risk for bio/pharma R&D supply chains is the considerable product development times that delay new product market launches (Griffiths, 2014); this risk extends to later development stages and aftermarket launch. It is not uncommon for a new product entering the third phase of clinical trials to fail to enter the market (Paul et al, 2010). These risks increase early-stage R&D investment costs (Herper, 2017), representing yet another barrier to success for smaller firms.

Just prior to the COVID-19 outbreak, the Disaster Recovery Institute International (DRI) released its Sixth Annual International Global Risk and Resilience Trends Report (DRI, 2020). The report identified significant threats and vulnerabilities that members believed would impact organizations in 2020 and beyond. The DRI survey participants represented a broad range of industries and jobs. This is a valuable list for bio/pharma firms including startups, because there are risks that are not specifically linked to specific industries. The list of emerging risks included pandemics as a significant threat.

METHODOLOGY

Our analysis consisted of an examination of business continuity practices across bio/pharma firms during the COVID-19 pandemic. We explored specific factors associated with pandemic impacts on R&D supply chains. Our focus was on determining how startup firms’ practices differed from mature firms regarding their implementation of a sales and operation plan, business continuity plan, and pandemic preparedness plan. We also evaluated whether startups’ R&D supply chains were affected differently relative to business process impacts, workforce impacts, and supplier impacts.

Two surveys were administered from July 2020 through February 2021. The first survey included questions primarily focused on work environments, although some questions asked about the existence of a BCP and COVID-19 effects on business processes. The second survey included questions more directly focused on pandemic-related planning and more detailed questions regarding effects on business processes. Participants were instructed to complete only those questions for which they had detailed knowledge. Table 1 includes the topics covered after combining aspects of both surveys relevant to this article.

Table 1: Survey Questions and Response Formats

No.	Question	Format
1	Did your organization have a pre-COVID SOP plan?	Yes or No
2	Did your organization have a pre-COVID BCP?	Yes or No
3	Did your organization have a pre-COVID pandemic plan?	Yes or No
4	Did you outsource 50% or more of supply chain activities?	Yes or No
5	How effective was your pandemic plan?	Rating
6	COVID effect on supplier's contractual obligations	Rate & Describe
7	COVID effect on your division's R&D commitments	Rate & Describe
8	COVID effect on your division's clinical trial schedules	Rate & Describe
9	COVID effect on your division's timelines for FDA approvals	Rate & Describe
10	COVID effect on your workforce's operational performance	Rate & Describe
11	COVID effect on your division's talent management capabilities	Rate & Describe
12	COVID effect on R&D suppliers’ lead times	Rate & Describe
13	COVID effect on R&D suppliers’ quality standards	Rate & Describe
14	COVID effect on R&D suppliers' cost structure	Rate & Describe

Many of the survey participants were members of professional groups that were located in or around Boston (Massachusetts), where 18 of the 20 largest bio/pharma firms have a presence. In addition, Boston and nearby Cambridge rank first in National Institutes for Health funding, venture capital funding, and laboratory space (Massachusetts Biotechnology Council, 2020).

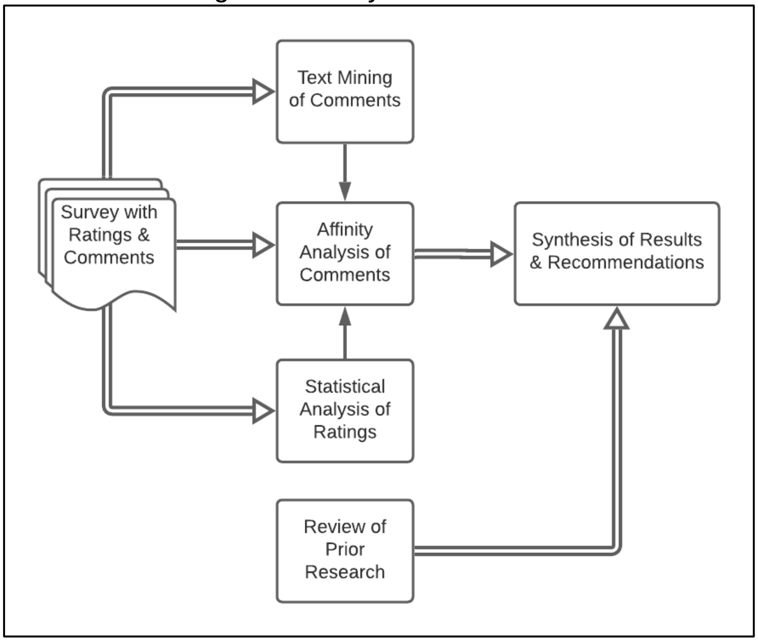
In addition, survey participants were asked the following:

1. What is the name of your firm and the division within which you work?
2. Would you classify your organization as a startup, transitioning, or mature?
3. Are you located in the U.S.? If so, in what State? If not, in what Country?
4. Would you classify your organization as U.S.-based or multi-national?
5. What is the primary business for which your organization operates?

The analysis of survey results applied a variety of qualitative and quantitative analytical methods, using the framework described in Figure 1. The analysis methodology extracted and categorized information while providing estimates of magnitude associated with the various categories. The three main types of analysis methodologies were affinity analysis, statistical analysis (including chi-square test for association), and text mining analysis (including sentiment analysis).

Affinity approaches were used to consolidate disparate information into categories without pre-determining the categories (Northcutt & McCoy, 2004, P. 95). For each qualitative survey question, the authors summarized each comment using consistent terminology (e.g., “Zoom,” “Microsoft teams” and “Teleconferencing” all refer to a means of communicating remotely). Then, each summarized statement was combined into a common category, based on summarized comments having a common focus, intention, or meaning. The benefit of this approach is that categories are not defined in advance, minimizing analyst bias. Visualizations were created that convey the frequency of responses for each qualitative category determined by the affinity approach.

Figure 1: Analysis Framework



MINITAB software was used to perform the statistical analyses for quantitative responses. Because most responses were categorical, the main method employed was to determine significant differences across respondents' characteristics, such as their job title, years of experience, location, and company status. Chi-square tests for association were used for this analysis, where some categories were combined to achieve a suitable sample size in each category (e.g., no impact and minimal impact were often combined). In the results section below, some p-values are provided, but in all cases significance was declared when the p-value fell below 0.05 (a p-value conveys the statistical likelihood that results are random).

Text mining was also employed for analyzing the qualitative (i.e., free text) information. Free text responses can provide more diverse explanations of respondents' experience than the numerical counterparts. For example, free text responses can help explain why certain nominal choices were made (Jackson & Trochim, 2002). Text mining has been successfully applied in many environments, including criminal detection and social media content (Danneman & Heimann, 2014, P. 7). This method is not affected by the human bias that can accompany affinity or other manual approaches to evaluating surveys (Roberts et al, 2014). Test mining is less time consuming than manual methods (Grimmer & Stewart, 2013), and its routines do not suffer from human bias or inconsistency (Chai, 2019). The text mining methods applied here were word frequency analysis (Shmueli et al, 2020, p. 76) and sentiment analysis (Silge & Robinson, 2017, P. 16).

RESULTS

In total, 320 participants completed one of the surveys; they were split about equally among the first and second surveys. The breakdown of firms was 19% startup, 20% transitioning, and 61% mature. Sixty percent of the firms represented in the survey were multi-national. Of the participants completing the survey, 78% were located in the U.S. The analysis results are detailed below.

Effectiveness of Business Continuity and Pandemic Plans

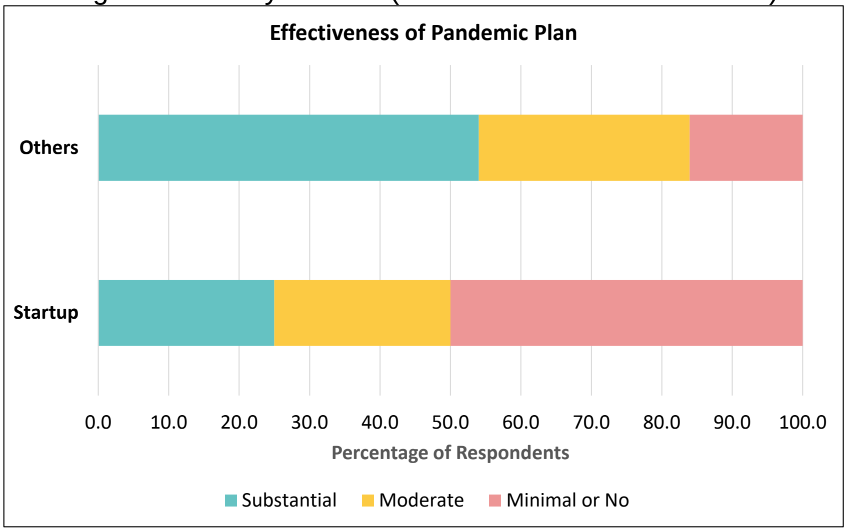
Table 2 shows the results for the four yes/no survey questions. The differences across firm status were statistically significant for questions 1-3 (with p-values 0.000, 0.003, and 0.012 respectively). Compared to mature firms, startups were less likely to have a sales and operations planning (SOP), they were less likely to have a BCP, and they were less likely to have a pandemic plan. The differences in the level of outsourcing across firm status was were not statistically significant.

Table 2: Survey Results (Yes/No Questions)

Question: Did your organization….	**Startup**	**Transitioning**	**Mature**
1. …have a pre-COVID SOP plan?	24%	29%	87%
2. …have a pre-COVID BCP?	38%	55%	80%
3. …have a pre-COVID pandemic plan?	11%	21%	48%
4. …outsource 50% or more of supply chain activities?	40%	63%	44%

Results for the key question concerning the effectiveness of their firm's pandemic plan (Question 5) are displayed in Figure 2 (where "Others" include mature and transitioning firms). Pandemic plans for startups were significantly less effective than those for transitioning or mature firms (p=0.001). It is interesting that firms self-classified as transitioning were less likely to have a BCP or pandemic plan in place, but those that did had plans that did not differ in effectiveness from mature firms.

Figure 2: Survey Results (Pandemic Plan Effectiveness)



In summary, survey results indicate that mature bio/pharma companies were better prepared and positioned to deal with the impacts of COVID-19. The fact that so many startups in the sample have not formally addressed business continuity and pandemic preparedness (although not entirely unexpected) represents a substantial weakness. It also presents an opportunity to provide assistance to help startups address business continuity planning needs going forward.

Effects on R&D Activities: Business Processes

Questions 6-9 concern impacts on business processes, with results summarized in Table 3. The most prominent impact was on clinical trial schedules. Impacts on other business processes (contractual obligations, R&D commitments, and FDA appraisals) were less evident, albeit affected at least moderately by about 50% of the respondents. Although there may be a tendency for more impacts at startup firms, none of the effects on business processes were statistically significant across firm status.

Table 3: Survey Results (Effects on Business Processes)

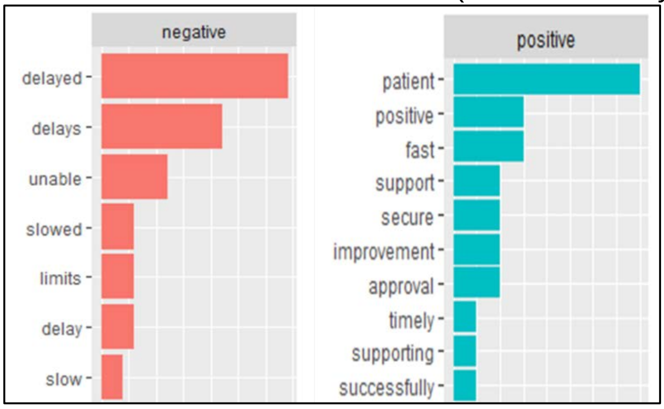
Question: Rate the COVID-19 effect on your...	Minimal	Moderate	Substantial
6. ...contractual obligations.	51%	32%	18%
7. ...R&D commitments.	45%	38%	16%
8. ...clinical trial schedules.	28%	34%	38%
9. ...timelines for FDA approvals.	57%	21%	21%

Text mining approaches were also applied to the total of 2471 words across the two surveys in response to open-ended questions involving this topic. The word cloud in Figure 3 shows the results of the word frequency analysis, which are consistent with those in Table 3. The most frequently used words with the most relevant meanings included: trial, delay, clinical, study, and patient.

Figure 3: Effect on Business Processes (Word Cloud)

Sentiment analysis was performed to supplement the word frequency analysis, with the results shown in Figure 4. The methodology used a standard dictionary to associate positive and negative sentiments with each text response. Although this method is compromised by the ambiguous sentiment associated with some words (e.g., patient), it highlighted many comments on clinical trials and delays associated with their completion. Overall, the total weighted sentiment score was negative.

Figure 4: Effect on Business Processes (Sentiment Analysis)



In summary, the pandemic impacted many aspects of business process performance. The most frequently cited impact was on clinical trial schedules. The impacts on clinical trials were mainly associated with delays.

Effects on R&D Activities: Workforce

Questions 10 and 11 concern impacts on work performance and talent management, with the results summarized in Table 4 (where H=Highly, M=Moderately, N=Negative, and P=Positive). The overall tendency was a balance of positive and negative impacts (the marginal tendency in the negative direction is not statistically significant). Although the tendency was for more negative impacts at startup firms, the differences were not statistically significant.

Table 4: Survey Results (Effects on Workforce)

Question: Rate the COVID-19 effect on ...	**HN**	**MN**	**None**	**MP**	**HP**
10. ...workers' operational performance.	8%	29%	30%	20%	13%
11. ...talent management capabilities.	11%	24%	38%	16%	11%

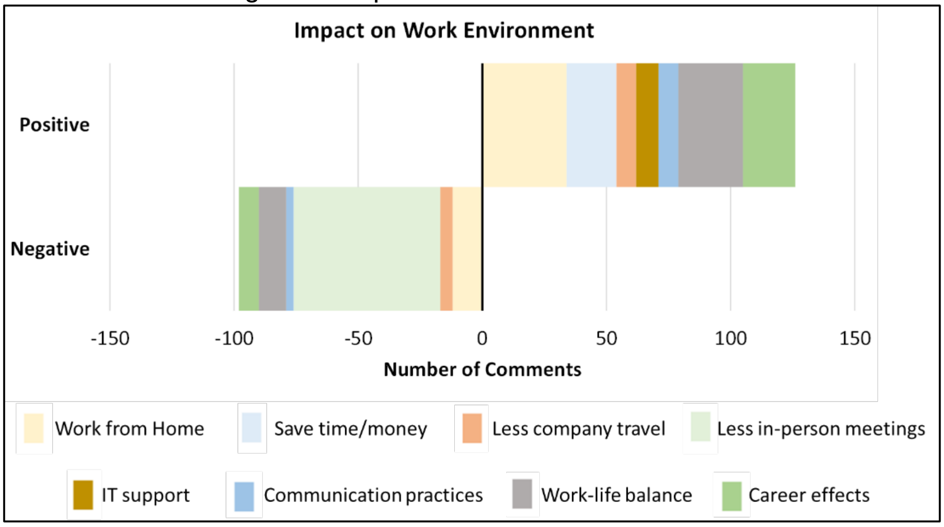
Regarding work environments, text mining was applied to the total of 10459 words across the two surveys that were submitted in response to open-ended questions concerning this topic. The word frequency results are not shown because their frequencies were dominated by words associated with work-from-home, such as "remote" and "home." The result of the sentiment analysis is shown in Figure 5. The analysis results are interesting because positive sentiments outweighed negative sentiments. The display highlights the major positive impacts (flexibility, benefits) and the main negative impacts (difficult, hard, stress).

Affinity analysis was also performed for the work environment text questions, with the results shown in Figure 6. Here, categories were created that could represent either a positive or negative impact depending on the author's sentiment. The number of positive comments was about 28% higher than the negative comments. The most cited negative comment was less in-person meetings. The three most cited positive comments were: work-from-home, work-life balance, and career effects. Younger workers reported more positive effects than their older, more experienced counterparts. The affinity approach was also applied to questions on the survey regarding operational impacts on workers' jobs. The most widely cited impact was reduced capacity, with specific details that included offices and laboratories being closed, operations and laboratory experiments being halted or delayed, personnel recruitment halted, projects delayed, longer lead times, and reduced operational efficiencies.

Figure 5: Effect on Work Environment Sentiment Analysis

In summary, impacts on operational performance of workers and talent management practices were moderately negative. Impacts on work environments were somewhat positive, with younger workers citing many advantages of work-from-home, but more mature workers concerned about fewer in-person meetings.

Figure 6: Impacts on Work Environment



Effects on R&D Activities: Suppliers

Questions 12-14 concern impacts on suppliers' operations, with results summarized in Table 5. The three questions were designed to correspond to the "better, cheaper, faster" (BCF) operational performance goals. Companies were better able to maintain quality levels (i.e., better) and cost structures (i.e., cheaper) as compared to a poorer ability to maintain schedules (i.e., faster). About two-thirds of the respondents reported moderate or substantial impacts on lead times. Although the overall tendency was for more negative impacts at startup firms, the differences were not statistically significant.

Text mining approaches were applied to the total of 4504 words across the two surveys that were submitted in response to open-ended text for questions involving this topic. The word frequency results are not shown because they were dominated by words similar to "communicate" (and others with this root or similar meaning). The sentiment analysis had an overall negative sentiment, and included somewhat ambiguous but negative sentiments such as "issue," "difficult," and "trouble."

Table 5: Survey Results (Effects on Suppliers)

Question: Rate the COVID-19 effect on ...	Minimal	Moderate	Substantial
12. ...R&D supplier's lead times.	33%	39%	26%
13. ...R&D suppliers quality standards.	68%	21%	11%
14. ...R&D suppliers cost structure.	48%	42%	10%

An affinity analysis was performed on the text responses that concerned impacts on suppliers, in combination with an additional question from the first survey concerning impacts on customers. The top 6 categories of impacts were: office closures (including laboratories), travel restrictions (including pickups and deliveries), delivery delays (including increased turnaround times), communication difficulties (including getting answers quickly), program or project delays (including clinical trials), and supply shortages (including important commodities). These results also reinforced the tendency for impacts causing delays.

In summary, survey results indicated that, although all aspects of supplier performance were impacted, the biggest impact was on lead times. This result is consistent with the analysis of business processes that showed impactful delays for clinical trial schedules. Although any firm would be concerned about these delays, startups would be more vulnerable because of their lack of significant financial resources. Startups would be well-served by developing business continuity plans that leverage their rapid decision making to mitigate this risk.

DISCUSSION AND RECOMMENDATIONS

The survey results indicated that most bio/pharma startups did not have a BCP or a pandemic plan in place prior to 2020, and that startups' business continuity plans were less effective than those within mature firms. In addition, the surveys revealed that, during the COVID-19 pandemic, the most impactful effect on bio/pharma firms was delays – especially clinical trial schedule delays. Smaller impacts were reported on quality and cost structures. The effect on workplace management was moderately negative although the impact on work environments was somewhat positive, especially for younger workers.

The literature analysis indicated that startups have fewer resources that can be devoted to supporting a BCP, including financial and knowledge-related resources. Therefore, startup leadership has competing priorities and the development of a BCP may not appear worthy of attention. But, startups also have advantages. Although lacking the resources of large firms, startups' organizations tend to be flexible because of their simplified management structure and direct operational involvement of C-level executives. These characteristics result in rapid decision making with fewer bureaucratic obstacles.

Companies of all sizes have to account for a wide variety of business disruptions. A useful approach is based on the incident spectrum (Banasiewicz & Weidman, 2016, p. 205-206). The four stages of the incident spectrum include outages (e.g., loss of power for several hours), significant events (e.g., IT system down for several days), disasters (e.g., fire causing inability to conduct business for several weeks), or catastrophic events (e.g., the COVID-19 pandemic whose disruption was evident for over one year).

Designing, implementing and maintaining an effective BCP can be a resource intensive and time sensitive undertaking. Although larger or mature firms can devote resources to this endeavor, startups do not have the capital to invest in developing a formal and comprehensive BCP. For example, a mature firm often includes a function devoted to business continuity management reporting to a C-level executive. Startups would need to satisfy the core requirements of a BCP within the parameters of their organizational structure.

Based on the state of knowledge and current practices in business continuity management, we recommend a list of five fundamental and cost-effective steps that startups should take in developing a basic BCP. To provide perspective, the recommendations are discussed in the context of both mature firms and startups. Although the focus of this article is on bio/pharma, the recommendations below would apply to any industry.

1. Establish a Crisis Management Team

A crisis management team (CMT) is responsible for directing a company's response and recovery efforts through its command control and communication activities. The CMT is necessary to minimize the damage and chaos associated with a disruptive event. Large firms will typically have a centralized CMT supported by a number of subordinate teams based on various functions, technologies, or locations. The total number of CMT members across these firm would total 20 to 60 members. A startup will likely have one CMT, while transitioning firms may implement a decentralized CMT structure with one CMT for each geographic location or specialized business area. For startups with one CMT, leadership needs to ensure that it is a flexible decision-making body that is capable of dealing with a wide array of events as identified in the incident spectrum. The CMT for a startup may consist of 5 or 6 members, including members familiar with information technology, human resources, facilities and safety management, and business operations.

2. Perform a Risk Assessment and a Business Impact Analysis

The risk assessment (RA) provides a clear picture of an organization's threats and vulnerabilities (i.e., risks), which can be substantial and will expand over time. It lists critical risks and, for each, it estimates likelihood, quantifies impacts, and identifies mitigation strategies. The RA addresses both risks which the firm has control over as well as risks that the firm has only indirect or no control. A business impact analysis (BIA) identifies each business disruption's quantitative and qualitative impact for the company's critical business functions, The BIA includes recovery priorities and interdependencies to establish recovery time objectives, recovery point objectives, and minimum resource requirements. These components should be verified and tested periodically.

The conclusions reached through the RA and BIA provide an actionable basis for developing continuity strategies and business continuity plans. Larger firms will expend considerable resources in performing a periodic RA and BIA, often utilizing third-party software applications (and/or consulting firms) that can facilitate their completion. A less costly option for a startup is having the RA and BIA performed by members of the CMT. This group is familiar with the company's operations and the goal of the BCP. Startups can also use standard RA and BIA documentation templates that are freely available in the public domain, such as the Hazards Vulnerability Analysis Tool developed by Kaiser Permanente (2021).

3. Identify and Implement Business Continuity Strategies

The RA and BIA should provide the information needed to identify recovery strategies for the firm's operations and technology. Each strategy should be evaluated for plausibility using criteria such as functionality, cost-benefit, timing considerations, existing mitigation plans, and current control measures. Recover strategies should be flexible and support the business-based recovery time objectives of the firm. The ideal recovery strategy vision would be the creation of an alternative operating environment that can be implemented immediately after the disruptive event.

Communication is essential during times of crisis. Poor communication can exacerbate other impacts including schedule delays and quality issues by postponing their reconciliation. It is important for startups to address communication issues so that they can take advantage of their rapid decision-making tendency. Large firms will use a variety of communication strategies, including website updates, emergency notification systems, and crisis management applications. Work-from-home capabilities and data management systems (e.g., to prevent cyber-attacks) are also critical components of the business continuity strategy.

The best approach for the development of recover strategies and communication approaches for a startup would be commensurate with their size and financial resources. The approach should include a combination of several options, including: (a) a single, focused crisis communication strategy based on text messaging, emails, and employee call trees; (b) remote access capabilities and telecommunication systems that can support work-from-home; (c) basic data backup and records management programs including both electronic and hard copy data; (d) use of manual processing and reciprocal agreements with key vendors and suppliers; (e) sharing knowledge and resources with other startups through industry and professional groups; and (f) mitigating controls associated with the facility's infrastructure and IT systems such as power generators, internet accessibility, and data center redundancies (which can often be negotiated in office leasing agreements).

4. Develop a Business Continuity Plan

The business continuity plan contains the processes and procedures necessary to guide the company through its response and recovery operations. The plan should include coverage of the three components of a corporate BCP: (a) incident response and management, (b) technology recovery and business recovery, and (c) resumption. For a startup with a single CMT, a single business continuity plan would be developed. It should be designed to cover a wide range of incidents, and include worst-case scenarios. In addition, it should be actionable and easy to understand, and it should be precise in describing business recovery activities and associated protocols. Finally, it should be flexible so that it captures ongoing business environment changes.

Several third-party applications exist for the development and retention of business continuity plans. These applications are more appropriate for larger firms because of their implementation costs and ongoing maintenance requirements. When a startup develops a business continuity plan, priority should be given to timely decision-making and the plan's detail should not be burdensome. Their plan's documentation would be better accomplished using widely available word-processing software applications rather than third-party applications.

5. Integrate with Supply Chain

Because so many R&D activities are outsourced, the impact of supplier disruptions cannot be ignored or downplayed. The extended enterprise includes vendors and suppliers, as well as public utilities, financial institutions, internet/telecom service providers, and logistics providers. Emphasis should be placed on integrating a firm's business continuity strategies and recovery planning with those of its key suppliers and business partners. Mature firms can take advantage of their leverage to create mutually-beneficial business continuity processes within contracts and service-level agreements (SLAs). Startups will need to include the extended supply chain in their BCP but without significant collaboration with suppliers and vendors.

A good gauge of integration effectiveness would be the estimation of recovery times. Recovery time estimates can be overly optimistic because, in many cases, recovery occurs as fast as the slowest supply chain member. The extended enterprise can be integrated in a startup company's planning are by obtaining copies of incident response and business continuity plans, or by participating in joint exercises.

Summary

Table 6 summarizes the differences in how the five recommendations listed above can be implemented in a startup firm. It provides perspective by comparing aspects of the startup approach to the approach typically undertaken by a mature firm.

Table 6: Summary of BCP Modifications for Startups

Mature Firm Approach	**Startup Approach**
Formal BCP department or function.	Shared BCP responsibilities.
Multiple CMTs with 20-60 members.	Single CMT with 4-6 members.
BIA for all business functions.	BIA for critical business functions.
Third party RA and BIA software.	Public domain templates.
RAs and BIAs performed by consultants.	RAs and BIAs done by the CMT.
Workforce mobility program.	Remote working capability.
Multi-layered communication strategy.	Single communication strategy.
Cutting edge data management & backup.	Basic data backup.
High availability recovery strategies.	Manual processing/reciprocal agreements.
Knowledge exists or can be acquired.	Knowledge sharing in professional groups.
Advanced facilities & IT mitigation controls.	Basic facilities & IT mitigation controls.
Multiple business continuity plans.	Single business continuity plan.
Third party BC planning software.	Self-generated & maintained files.
Contractual leverage & SLAs.	SCM included as part of the BCP.

FUTURE WORK

The COVID-19 pandemic provided a useful laboratory for studying business continuity management at bio/pharma firms. Surveys of professionals within these firms highlighted the need for better business continuity planning at startups, which have fewer resources but also faster decision making tendencies. Recommendations were made for startups using a five-step approach. Future work will include follow-ups with the survey participants who plan to implement some or all of the recommendations. Their progress will be evaluated and reported in future publications.

REFERENCES

Alarape, A.A. (2007). Entrepreneurship programs, operational efficiency and growth of small businesses. *Journal of Enterprising Communities: People and Places in the Global Economy, 1*(3): 222–239.

Alshawi, S., Saez-Pujol, I., & Irani, Z. (2003). Data warehousing in decision support for pharmaceutical R&D supply chain. *International Journal of Information Management*, *23*(3), 259-268.

Altman, E., A., Sabato, G., & Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *Journal of Credit Risk, 6*(2), 95.

Atlas, L., & Sobotka, P. (2013). Third-party vendor management: considerations for third-party management in a risk-focused environment. *Applied Clinical Trials, 22*(9), 18-20.

Andersson, T., Gleadle, P., Haslam, C., & Tsitsianis, N. (2010). Bio-pharma: A financialized business model. *Critical Perspectives on Accounting, 21*(7), 631–641.

Ayati, N., Saiyarsarai, P., & Nikfar, S. (2020). Short and long term impacts of COVID-19 on the pharmaceutical sector. *Daru, 28*(2), 799–805.

Banasiewicz, A.D. & Weidman, D.A. (2016). *Total Exposure Management: Risk, Resilience, Change*. Bristol: Erudite Analytics.

Business Continuity Institute (2017). *Glossary of Business Continuity Term*s. https://www.thebci.org/resource/bci---drj-glossary-of-terms.html.

Campbell, J.J. (2018). *Understanding Pharma 2/E*. Raleigh, NC: Pharmaceutical Institute.

Chai, C.P. (2019). Text mining in survey data. *Survey Practice, 12*(1), 1-14.

Chong, S. (2014). Business process management for SMEs: An exploratory study of implementation factors for the Australian wine industry. *Journal of Information Systems and Small Business 1*(1/2): 41–58.

Classen, N., Carree, M., Van Gils, A. & Peters, B. (2014). Innovation in family and non-family SMEs: An exploratory analysis. *Small Business Economics 42*(3): 595–609

Cohen, M.J. (2020). Does the COVID-19 outbreak mark the onset of a sustainable consumption transition? *Sustainability: Science, Practice and Policy, 16*(1), 1-3.

Cowling, M., Brown, R., & Rocha, A. (2020). Did you save some cash for a rainy COVID-19 day? The crisis and SMEs. *International Small Business Journal, 38*(7), 593–604.

Danneman, N. & Heimann, R. (2014). *Social Media Mining with R*. Olton: Packt Publishing, Limited.

Dahlberg, R., & Guay, F. (2015). Creating resilient SMEs: Is business continuity management the answer? *Sustainable Development, 2*, 975- 984.

Deloitte (2013). *Pursuing Risk Intelligence in a Rapidly Changing Industry*. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Governance-Risk-Compliance/dttl-grc-pursuingriskintelligenceinrapidlychangingindustry.pdf.

DiMasi, J.A., Hansen, R.W., & Grabowski, H.G. (2003). The price of innovation: New estimates of drug development costs. *Journal of Health Economics, 22*(2), 151-185.

DRI (2020), Disaster Recovery Institute International: Sixth Annual International Global Risk and Resilience Trends Report.

Du Toit, G.S., Erasmus, B.J. and Strydom, J.W. (2009). *Definition of a Small Business and Introduction to Business Management. 7/E*. Cape Town: Oxford University Press.

Edvardsson, I.R., Durst, S., & Oskarsson, G.K. (2019). Strategic outsourcing in SMEs. *Journal of Small Business and Enterprise Development, 27*(1), 73–84.

Falkner, E.M., & Heibl, M.R.W. (2015). Risk management in SMEs: a systematic review of available evidence. *The Journal of Risk Finance, 16*(2), 122–144.

Griffiths, J. (2014). High value, high uncertainty: Measuring risk in biopharmaceutical research and other industries Investing in the future of health. Deloitte. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-measuring-risk-in-biopharmaceutical-research.pdf.
Grimmer, J., & Stewart, B.M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267-297.
Grover, A., & Suominen, K. (2014). 2014 Summary – State of SME Finance in the United States. *TradeUp (January)*, 1–38.
Haase, A., & Eberl, P. (2019). The challenges of routinizing for building resilient startups. *Journal of Small Business Management, 57*(sup2), 579–597.
Handfield, R., Finkenstadt, D.J., Schneller, E.S., Godfrey, A.B., & Guinto, P. (2020). A commons for a supply chain in the post-covid-19 era: The case for a reformed strategic national stockpile. *The Milbank Quarterly, 98*(4), 1058–1090.
Herper, M. (2016, March 24). The cost of creating a new drug now $5 Billion, pushing big pharma to change. Forbes. Retrieved April 24, 2021, from https://www.forbes.com/sites/matthewherper/2013/08/11/how-the-staggering-cost-of-inventing-new-drugs-is-shaping-the-future-of-medicine/?sh=182745e213c3.
Husin, M.A. and Ibrahim, M.D. (2014). The role of accounting services and impact on Small Medium Enterprises (SMEs) performance in manufacturing sector from east coast region of Malaysia: A conceptual paper. *Procedia-Social and Behavioral Sciences 115*, 54–67.
Jackson, K.M., & Trochim, W.M. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods, 5*(4), 307-336.
Kaiser Permanente (2021). Updated HVA Tool. Retrieved from: https://www.calhospitalprepare.org/post/updated-hva-tool-kaiser-permanente.
Kilpatrick, J., & Bater, L. (2020). COVID-19: Managing supply chain risk and disruption. Deloitte. https://www2.deloitte.com/global/en/pages/risk/cyber-strategic-risk/articles/covid-19-managing-supply-chain-risk-and-disruption.html.
Kumar, A., Luthra, S., Mangla, S.K., & Kazançoğlu, Y. (2020). COVID-19 impact on sustainable production and operations management. *Sustainable Operations and Computers, 1*, 1-7.
Maini, F. (2014). Risk and regulation: A wave of change for pharma. Deloitte. Retrieved April 23, 2021, from https://blogs.deloitte.co.uk/health/2014/07/risk-and-regulation-a-wave-of-change-for-pharma.html
Maurya, U.K., Mishra, P., Anand, S. and Kumar, N. (2015). Corporate identity, customer orientation and performance of SMEs: Exploring the linkages. *IIMB Management Review 27*(3): 159–174.
Murphy, W.H. (2016). Small and mid-sized enterprises (SMEs) quality management (QM) research (1990-2014): a revealing look at QM's vital role in making SMEs stronger. *Journal of Small Business and Entrepreneurship, 28*(5), 345–360.
Northcutt, N., & McCoy, D. (2004). *Interactive Qualitative Analysis: A Systems Method for Qualitative Research*. Portland: Ringgold, Inc.
Lucky, E.O., & Olusegun, A.I. (2012). Is small and medium enterprises (SMEs) an entrepreneurship? *International Journal of Academic Research in Business and Social Sciences 2*(1): 487-497.
Paul, S. M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., & Schacht, A.L. (2010). How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery, 9*(3), 203-214.
Polishchuk, V. et al. (2019). A Fuzzy Model of Risk Assessment for Environmental Start-up Projects in the Air Transport Sector. *International Journal of Environmental Research and Public Health, 16*(19), 3573.

Roehr, B. (2020). Bringing drug production home: how the US is rebuilding the drug supply chain after COVID-19. *BMJ, 370*, m3393–m3393.

Socal, M.P., Sharfstein, J.M., & Greene, J.A. (2020). Critical drugs for critical care: Protecting the US pharmaceutical supply in a time of crisis. *American Journal of Public Health, 110*(9), 1346–1347.

Shmueli, G., Bruce P.C., Gedeck P. and Patel, N.R. (2020), *Data Mining for Business Analytics: Concepts, Techniques and Applications in Python*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Shackelford, B. & Moris, F. (2016). A snapshot of business R&D employment in the United States. National Science Foundation. Retrieved from https://www.nsf.gov/statistics/2017/nsf17302/

Shehab, M.A. (2008). Factors influencing accounting information system performance among small and medium enterprises (SMEs) in Tropli. MSc thesis. Universiti Utara Malaysia.

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. Sebastopol, CA: O'Reilly.

Sullivan-Taylor, B., & Branicki, L. (2011). Creating resilient SMEs: Why one size does not fit all. *International Journal of Production Research, 49*(18), 5565-5579.

Van Norman, G.A. (2016). Drugs and devices: Comparison of European and U.S. approval processes. *JACC: Basic to Translational Science, 1*(5), 399-412.

Wagner, E., & Burgstaller, J. (2015). How do family ownership and founder management affect capital structure decisions and adjustment of SMEs? *The Journal of Risk Finance, 16*(1), 73–101.

Yarza, C. (2019). Managing regulatory compliance. PwC. Retrieved April 23, 2021, from https://www.pwc.com/il/en/pharmaceuticals/managing-regulatory-compliance.html.

DECISION SCIENCES INSTITUTE

Design and Analysis of a Cloud-based Fleet Operation Management Platform with Smart Sensing and Predictive Maintenance Integration

Yiyun (Cindy) Fei
Georgia Institute of Technology
Email: yfei38@gatech.edu

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Jianyuan Peng
Georgia Institute of Technology
Email: pjy@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

This paper proposed an integrated platform for airport GSE management using real-time data. A non-invasive smart sensor using a hydrostatic pressure transducer has been developed for data collection on the ground and data transmission to the cloud server. Recurrent neural network models would be used for maintenance prediction and planning. This paper also discusses the use of linear programming models for finding optimal solutions for flexible shift planning and vehicle route planning. The benefits of the proposed flatform include lower operation costs, higher resource utilization, and better communication between airport managers and ground staff.

KEYWORDS: Smart sensing, Predictive maintenance, Fleet operation planning, Airport operation management, real-time

INTRODUCTION

Aircraft tug trucks are widely used to transport airplanes on the ground. The truck maintenance requirement is determined by the level of oil inside the truck engine. GSE technicians need to climb on top of the vehicle to reach the engine and insert the dipstick to check the oil level. According to the requirement from the Occupational Safety and Health Administration (OSHA), industry employers must provide fall protection for workers working at a level higher than four feet from the ground. The implementation of fall protection is both time-consuming and expensive. Also, it is hard to keep track of the oil level in each tug truck due to the uneven workload of the tugs when exerting immense force during a short time when pulling an airplane, which causes the diesel oil to be consumed faster than normal activities. To simplify the oil level detection process, a new type of smart sensor for measuring the oil level is proposed.

Along with the development of the smart sensor, there are three technical challenges developed during the designing of the entire GSE management system. First, since real-time data can be collected using the sensors and upload to the cloud for analysis, the maintenance schedules should be predicted in advance to avoid an unplanned mechanical breakdown that causes a waste of resources. Second, after obtaining the maintenance schedules, shift planning needs to be performed based on the workload at different time periods since the starting time and duration of each shift are flexible. The optimal planning solution has to balance the maximum and minimum maintenance demand. The third challenge is to optimize the vehicle routings for job dispatching. By finding the shortest paths for each vehicle to complete all task demands, the operational cost can be minimized. Furthermore, the real-time data update is also a part of the route planning in case of an emergency failure happen at a certain resource.

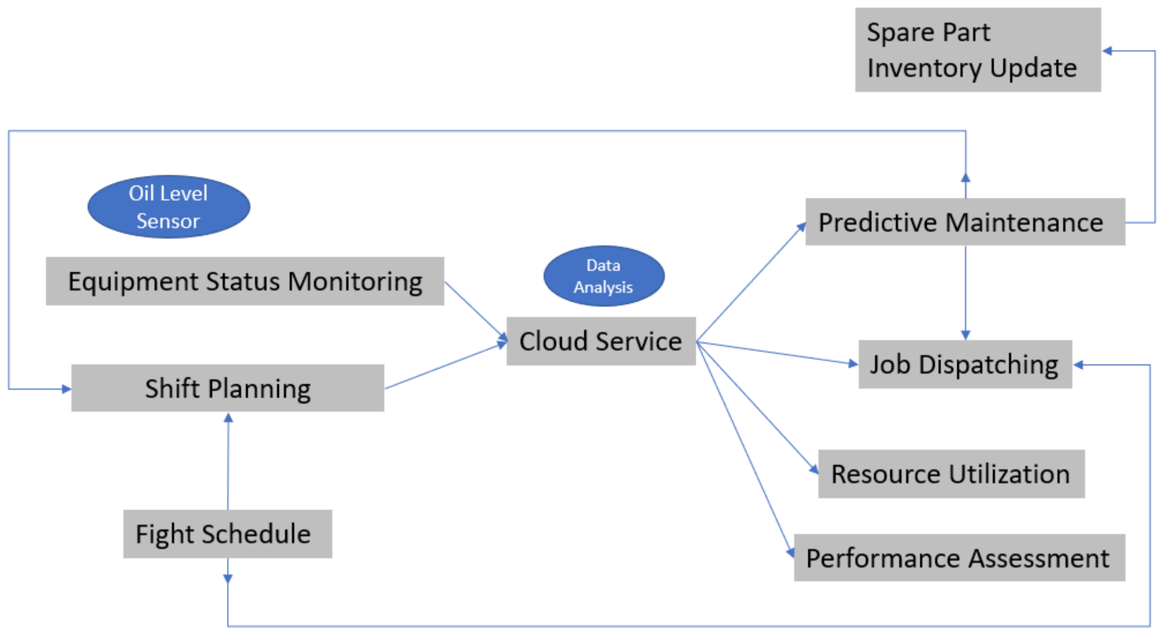
LITERATURE REVIEW

A lot of research has been done on the development of airport management improvements. Newbold had proposed the idea to transform the airport into a “real-time” digital twin, where the main components of the data-drive system include integrated planning and predictive maintenance. By using real-time data, better decisions could be made (Newbold, 2020). Legner and Thiesse use the application at Frankfurt airport to prove the feasibility of using real-time data as a part of the operational system and state that the benefits include better planning and control, structured documentation, and paperless information management (Legner & Thiesse, 2006). Recurrent neural networks (RNN) are used for predictive maintenance problems such as Remaining Useful Life (RUL) prediction and abnormal behavior detection (Rivas et al., 2019; Demidova, 2020; Arifoglu & Bouchachia, 2017). Predicted results for maintenance using different types of RNN, such as LSTM, GRU, and Vanilla have been compared (Arifoglu & Bouchachia, 2017; Ding et al. 2019; Rahhal & Abualnadi, 2020). A case study for using neural network models for airport GSE failure prediction has been done by Smith et al. (Smith et al., 2010). The fleet operation planning problem is essentially a vehicle routing problem, which can be solved using the mathematical method of linear programming. Wang et al. have proposed an optimal scheduling model of different types of GSE (Wang et al, 2020). The classic vehicle routing problem can be extended to include more constraints, such as specified time windows for arrival (Agra, et al., 2013).

SYSTEM DESIGN AND ANALYSIS

Figure 1 shows the proposed platform for operation management. The oil sensor and other types of equipment status monitoring sensors are the first input of this platform. The real-time data are collected and transmitted to the cloud server via Bluetooth and Wifi or LTE for analysis and further decision making. Fight schedules are another important input that determines the workload for shift planning and the time period required for completing each job at an airport. The four outputs from the cloud server are maintenance prediction, job dispatching, resource utilization, and performance assessment. Vehicle’s historical data and real-time data are used for predictive maintenance, where abnormal behavior would be detected, and the remaining useful life would be determined. Because maintenance is closely related to the spare part inventory, predicting the time and frequency of server needed also helps to manage the inventory. Once the updated spare parts inventory level is below the safety stock level, a request for reordering will be triggered. The estimated time period that server are required is also an input for optimizing shift planning and ground job dispatching. In case of an emergency vehicle breakdown, the job dispatching will be adjusted accordingly to provide an optimal solution at that moment. Because the status of each vehicle is being monitored, the overall performance and utilization of the vehicles can be organized and send to the management department for future operational improvement.

Figure 1: Structure of a cloud-based airport ground operation management platform



Smart Sensor Platform

In order to detect the oil level without requiring operators to climb above 4 feet, a non-invasive smart oil level sensor is developed. The smart sensor includes two components. The first component is the sensor hardware, which is a device that detects the oil level inside a vehicle's oil pan. According to the engineering requirements, the sensor design should be able to adapt to different engines or oil pan configurations so that it is universal and easy to install on various fleet vehicles. The sensor is expected to require minimal maintenance and provide accurate outputs. It also needs to withstand the heat of oil since it will be working both before and while the engine is in use. In addition, it also needs to adapt to a wide range of environmental temperatures because the tug that pulling back airplanes is working outside most of the time. The second component is a communication system between the sensor and the operators and other related staff. There will be indicating lights panel showing the oil level inside of the vehicle so that the driver knows the current oil level after turning on the ignition. And the oil level data will also be transmitted from the sensor to mobile devices by Bluetooth and then to cloud server for data analysis.

After conducting preliminary experiments, the hydrostatic pressure transducer is decided to be the more feasible solution for detecting oil levels. According to the hydrostatic equation, hydrostatic pressure is directly proportionate to the density and depth of the liquid. Figure 2 shows the oil pan. Because the oil pan always has drain plugs at the bottom, the idea of placing the pressure transducer at the drain plug is generated. The transducer will measure the change in pressure and output respective change in current. Arduino is used as a data acquisition device to connect the output data with indicating lights and mobile devices. Figure 3 shows the schematic of the connection, where a resistor is added to measure the voltage drop. According to Ohm's law, the voltage drop will be proportionate to the changing current induced by the transducer. Arduino uses the input voltage value and outputs to the LED lights and Bluefruit, which is a chip that sends data to mobile devices through Bluetooth.

Figure 2: Structure of the cloud-based airport ground operation management platform

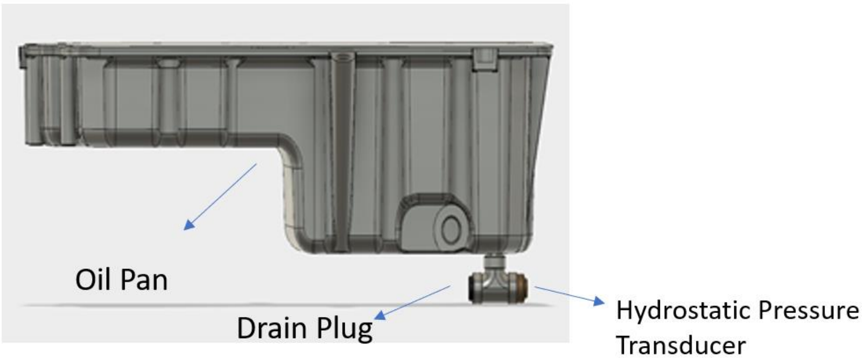
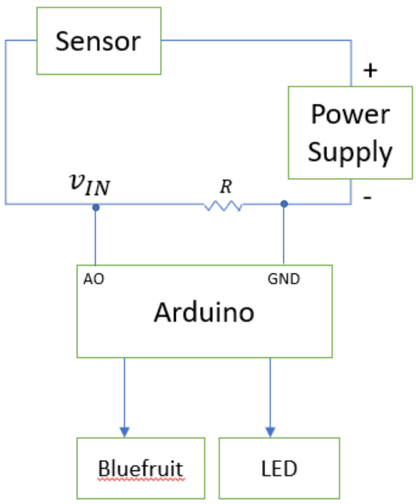


Figure 3: Sensor Connection Schematic



Subsequent experiments have been conducted with this design. First defining the acceptable oil level at a certain height, at any level lower than this threshold will be considered as unacceptable. Using the same oil that has been used in the resolution experiment, the result suggests that the pressure transducer can detect a slight change between acceptable and unacceptable levels. Linear regression has been applied to identify the relationship between output voltage $\boldsymbol{U}$ (V) and oil level $\boldsymbol{L}$ (inch), shown in Equation 1.

$$\boldsymbol{L = 13.8\,U + 0.0108} \tag{1}$$

Using this function, an exact real-time value of oil heigh can be calculated, rather than just identify “acceptable” or not. Figure 4 shows the indicating light design that will be mounted inside the vehicle. There are three LED lights, which are green, yellow, and red. The green light means the oil level is beyond 4 inches as the yellow light represents the oil level is between 2 and 4 inches. When the oil level drops below 2 inches, the red light will be turned on to notify the operator that server is required.

Figure 4: Indicating LED Showing three statuses

RECURRENT NEURAL NETWORK MODEL FOR PREDICTIVE MAINTENANCE

Predictive maintenance is a proactive method to forecast anomalies and potential failures of vehicles by monitoring the status of vehicles in operation. In order to reduce unplanned vehicle downtime and schedule maintenance in advance, real-time data is collected by sensors and analyzed by a recurrent neural network (RNN) model. The RNN model will be built and trained using vehicles’ historical operation data for initial predictions. Figure 5 shows the structure of the RNN model. The model’s inputs are data collected from different types of sensors, such as oil level sensors and tire pressure sensors. After feeding forward through several hidden layers, the predicted results will be available for further decision-making regarding the maintenance schedule.

Figure 6 is the process flow chart of the RNN for predictive maintenance. The model results will provide two types of predictions. The first type is remaining useful life prediction, which would be used for oil level sensors. By forecasting when the oil will drop below the minimum acceptable threshold, it is able to avoid potential damage on the vehicle caused by being short of oil. Thus, at what time the maintenance is needed can be determined and used for future optimal shift planning and job dispatching. Failure prediction is the second type of model output. This type of prediction can be used on equipment that runs continuously such as a motor or a pump, and anomaly detection will be performed by the RNN model. If abnormal behavior is detected, ground staff will be notified to check on the equipment and provide feedback on details about the failure. The expected time range to have server will be sent to shift planning or real-time job dispatching schedule as an input to determine the optimal time for server. In the case of a false alarm, where the vehicle is running normally but the model sends out a failure notification, the model will be retrained to adapt to the new pattern after receiving the feedback. As time passes, the model results would be more accurate.

Figure 5: RNN Model for Predictive Maintenance

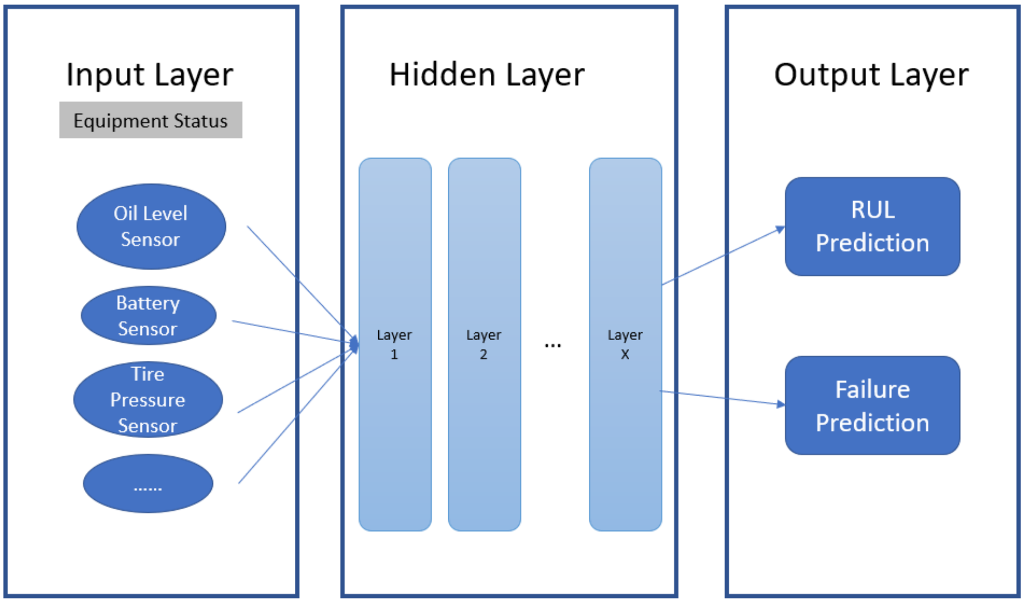
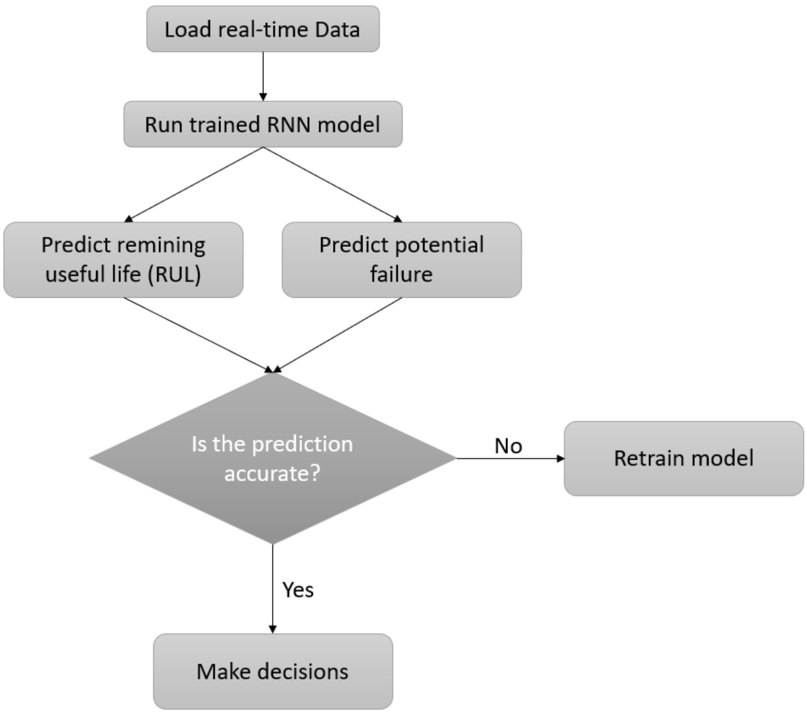


Figure 6: Process Flowchart for Deploying the RNN Model



SHIFT PLANNING FOR GROUND STAFF

Shift planning is a challenging decision-making process, whereby the goal is to determine the on-shift and off-shift periods of a day. As mentioned in the prior section, shift planning is generated based on the information provided from flight schedules and detected maintenance requirements. The number of incoming airplanes and the layover time until the next departure is the key factors that determine the workload, or demand. Since the starting time and duration of each shift are flexible, the ultimate solution for shift planning needs to select the optimal combination that minimizes the overall operation cost while avoid overstaffing or understaffing. However, an adequate amount of understaffing is considered to be acceptable for a special situation, such as for the night shift or the peak period, in order to reduce the more expensive cost due to overstaffing. Similarly, occasional overstaff for highly important jobs as a backup is also admissible. Therefore, the optimal solution balances the overstaff and understaff to meet the demand.

The method of mixed-integer programming is widely used for solving shift planning problems (Herbers, 2005; Di Gasperio et al., 2007). Equation 2 is the objective function where the goal is to minimize the sum of operation cost, $\boldsymbol{c}$, understaff cost, $\boldsymbol{u}$, and overstaff cost, $\boldsymbol{o}$, during a specific planning horizon. $\boldsymbol{S}$ represents a set of shifts in the planning horizon, $\boldsymbol{T}$. There are also different types of job requirements, $\boldsymbol{r}$, during the shifts. Equation 3-4 are the constrains that limits the shift design. Equation 3 states that if the number of staff in each shift must meet the demand, $\boldsymbol{D}$, after adding the understaff and/or subtract the overstaff. x and $\boldsymbol{m}$ are two binary variables indicating whether the shift is active at the time t and whether the number of staff in the shift is enough to meet the demand. Equation 4 restricts the overstaff and understaff cost and number of the shift, n, to be non-negative integers.

$$\mathbf{G} = \min \sum_{s \in S} c_s n_s + \sum_{t=1}^{T} \sum_{r \in R} u_{rt} + \sum_{t=1}^{T} \sum_{r \in R} o_{rt} \tag{2}$$

$$s.t. \sum_{s \in S} x_{st} n_s m_{sr} + \sum_{r \in R} u_{rt} - \sum_{r \in R} o_{rt} = D_{rt} \ \ \forall r, 1 \leq t \leq T \tag{3}$$

$$u_{rt}, o_{rt}, n_s \in \mathbb{N} \tag{4}$$

VEHICLE ROUTING OPTIMIZATION

Job dispatching includes dispatching of two types of resources: staff and vehicle. The dispatching of vehicles involves the classic problem of vehicle routing, where the goal is to find the most cost-effective paths for each vehicle to complete all the task demands. From the maintenance schedule, the vehicle availability will be known, and the time windows for different tasks will be given by the flight schedules. The paths for each vehicle can be both pre-determined before each shift and temporarily adjusted based on real-time feedback. For example, if there is an emergent mechanical breakdown, the vehicle routings will be recalculated to find the optimal solution in order to adapt to the current situation.

Equation 5 is the classic formulation for vehicle routing problems, where $\boldsymbol{C}$ is the cost for traveling from point i to point j, and $\boldsymbol{x}$ is a binary value 0 or 1 indicating whether this travel path for vehicle $\boldsymbol{k}$ is applied to the solution. The objective function is used to find the minimum operational cost for the vehicles in a set of $\boldsymbol{K}$ considering all possible locations. The maximal number of vehicles in a set of $\boldsymbol{K}$ depends on the maintenance schedule. Equation 6 is the first constraint. This constraint states that the arrival time at point j, $\boldsymbol{y_j}$, must be equal to or larger than the arrival time at point i, $\boldsymbol{y_i}$, plus the time for the vehicle to travel from position i to j. From the flight schedule and maintenance schedule, the time windows $[\boldsymbol{m_i}\ ,\boldsymbol{n_i}\ ]$ are known. The second constraint, Equation 7, shows that the arrival time at point i must be within the range of the provided time window.

$$\boldsymbol{Q} = \boldsymbol{min} \sum_{k=1}^{K} \sum_{i,j \in N} C_{ijk} x_{ijk} \tag{5}$$

$$y_j \geq y_i + \tau_{ijk} \tag{6}$$

$$m_i < y_i < n_i \tag{7}$$

CONCLUSIONS

The proposed airport operation management platform uses the real-time data collected from the smart sensors on different types of GSE and generates predictive maintenance schedules by using an RNN model. By implementing the smart sensors, the current problem which technicians need to set up fall protection each time checking the oil level while wasting a lot of time and money can be avoided. In addition, mathematical approaches such as linear programming are used to find the optimal solutions for shift planning and vehicle routings problems. This platform also provides statistical results such as resource utilization and performance assessment to the management department for further decisions of improvement. Further work will focus on the efficiency of the data analysis and field test with the proposed solution.

ACKNOWLEDGMENTS

Rayna Berinhout, Amber Fleeman, Jacob Jacob, Steven Maa, Laura Segura, and David Tsui Chang are acknowledged for their valuable contributions to prototype development and running the experiments.

REFERENCES

Agra, A., Christiansen, M., Figueiredo, R., Hvattum, L. M., Poss, M., & Requejo, C. (2013).*The robust vehicle routing problem with time windows*. Computers & Operations Research, 40(3), 856-866

Arifoglu, D., & Bouchachia, A. (2017). *Activity recognition and abnormal behaviour detection with recurrent neural networks*. 110, 86-93

Demidova, L. A. (2020). *Recurrent neural networks' configurations I the predictive maintenance problems*. IOP Conf. Ser.: Mater. Dci. Eng. 714, 012005

Di Gaspero, L., Gartner, J., Kortsarz, G., Musliu., N., Schaerf A., & Slany, W. (2007). The minimum shift design problem. Annals of operations research, 155(1), 79-105

Ding, H., Yang, L., & Yang, Z. (2019). *A predictive maintenance method for shearer key parts based on qualitative and quantitative analysis of monitoring data.* IEEE Access, 99, 1-1

Herbers, J. (2005). *Models and algorithms for ground staff scheduling on airports*. Bibliothek der RWTH Aachen.

Legner, C., & Thiesse, F. (2006). *RFID-based maintenance at Frankfurt airport.* IEEE Pervasive Computing, 5(1), 34-39

Newbold, A. (2020). *Transforming a functional airport to a smart, digital one*. Journal of Airport Management, 14(2), 106-114(9)

Rahhal, J. S., & Abualnadi, D. (2020). *IOT based predictive maintenance using LSTM RNN estimator.* 2020 ICECCE, 1-5

Rivas, A., Fraile, J. M., Chamoso, P., Gonzales-Briones, A., Sitton, I., & Corchado, J. M. (2019). *A predictive maintenance model using recurrent neural networks.* 14th International conference on soft computing models in industrial and environmental applications(SOCO 2019), 261-270

Smith, A. E., Coit, D. W., & Liang, Y.C. (2010). *Neural network models to anticipate failures of airport ground transportation vehicle doors.* IEEE transactions on automation science and engineering, 7(1), 183-188

Wang S., Che, Y., Zhao, H., & Lim, A. (2020). *Accurate Tracking, Collision detection, and optimal scheduling of airport ground support equipment.* IEEE Internet of Tings Journal, 8(1), 572-584

DECISION SCIENCES INSTITUTE

The automotive industry in San Luis Potosi, Mexico: a case of collaboration in clusters

Salvador T Porras
Universidad Autonoma Metropolitana-Iztapalapa
stp@xanum.uam.mx

Regina Leal Güemez
Universidad Autonoma Metropolitana-Iztapalapa
stp@xanum.uam.mx

ABSTRACT

Since Adam Smith *Wealth of Nations*, the traditional way to do business is to compete in open markets. More recently, in global markets, organizations seek other forms of conducting long-term economic activities based on collaboration including strategic alliances, business networks and clusters. In this paper we present an approach of clustering in the automotive industry located in the state of San Luis Potosi, Mexico. This strategy is an example of the importance of foreign direct investment (FDI) for developing a specific geographic area where different and diverse enterprises collaborate for achieving a common goal, for creating knowledge and achieving innovation.

KEYWORDS: Collaboration, clusters, Industry 4.0, San Luis Potosi, Circular Economy.

INTRODUCTION

The traditional way to conduct economic activity since Adam Smith's publication in 1776 (The Wealth of Nations) based on free trade is changing for a more collaborative way of doing business. Organizations seek long-term relationships for achieving their goals. Globalization has made organizations more and more interdependent on each other. The traditional short-term vision of conducting economic activity needs to change for a more long-term perspective based on collaboration (Park 1996). Collaboration is argued to be a necessity for organizations in order to compete in global economies (Astley 1984; Park 1996). Collaboration is a vital element for organizations for survival and development in a more and more turbulent environment (Emery & Trist 1965).

LITERATURE REVIEW

Collaboration allows organizations to achieve something in common that otherwise would not achieve (Axelrod 1984, Hamel, Doz & Prahalad 1989; Harrigan 1988; Ohmae 1989). Gray (1989: 5) considers collaboration as "a process through which parties who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible".

Organizations seek to combine their strengths with other organizations in order to overcome their own weaknesses. Some of the reasons for establishing collaboration exposed by organizations relate to access to new technologies or new markets, to fortify their position in the market, to achieve economies of scale, risk sharing, access to knowledge and to complement skills (Powell 1987). Research and development that conduct to innovate of new products has become a growing necessity. Collaborative arrangements could be particularly conducive to organizational learning or the transfer of organizational intangibles such as knowledge (Larson 1992; Larsson et al. 1998; Polanyi 1957)

Weber and Khademian (1997) identified three basic elements for initiation and maintenance of a collaborative effort: entrepreneurial leadership, organizational credibility, and a set of rules limiting any opportunistic behavior by participants of a collaborative arrangement.

The speed of communications using Internet, e-mail and other computer technologies has meant that changes immediately diffuse around the world, making product life cycles shorter. In these conditions, firms are increasingly obliged to face the challenges of technological change. Organizations are adopting “collective strategies” (Astley & Fombrun, 1983; Bresser, 1988; Bresser & Harl, 1986; Thorelli, 1986) to cope with rapid changes of technology.

The increase of global communications and trade has also provoked new patterns in the relationship between organizations. Collaboration is basic to these newly established relationships. As Gray (1989) proposes, we are in an era where collaboration, rather than competition, will govern the relationships between organizations.

Collaboration is a co-operative strategy that may also lead to competition, although a different form of competition (Hamel, Doz & Prahalad 1989). Important to say though that collaboration does not inhibit competition.

Why to collaborate?

Different are the reasons for organizations to establish cooperative arrangements with other organizations. Ring and Van de Ven (1992) state that among the most significant pressures for inducing collaboration are, rapid changes in technology, the competitive environment, and firm strategies. Porter (1985) suggests that firms that are pursuing a diverse set of objectives that involve reciprocal dependencies are more likely to collaborate. These objectives include: gaining access to new technologies or markets; benefiting from economies of scale in joint research, production, and marketing, and gaining complementary skills. Other advantages of collaborating include sharing the risks for activities that are beyond the scope or capability of a single organization, and gaining synergy by combining the strengths of firms, marketing joint venture, or technology licensing arrangement.

For collaborating, trust is important as a way to avoid opportunistic behavior (Simon 1947). Participants must trust in other participants. It is not completely or blind trust but a minimum level of trust is required. Qualities such as: integrity, loyalty, competence, consistency, and openness should be present for trusting (Gambetta, 1988; Ring & Van de Ven, 1994). Trust can reduce ‘transaction costs’ (Williamson, 1975, 1985) associated with exchanges with others. Trust is based on the belief in others behavior. Trust can be understood as “confidence in the other’s goodwill” (Ring & Van de Ven, 1992). All participants in a relationship should work for the common interest, not the individual (Porras, Clegg & Crawford 2004). Trust is not an asset that can be interchanged; it belongs to an actor.

Research suggests that different areas for collaboration are available to organizations, particularly to small and medium-sized enterprises (SMEs) (Larson & Starr 1993; Myhrvold et al. 1995; Powell 1987). Collaboration between organizations occurs when there is a partial overlap of their mission and domain (Buttery & Buttery 1994; Gray 1989; Thorelli 1986). Without the existence of partial overlap in scope, collaboration between organizations is unlikely to be achieved (Thorelli 1986).

Forms of collaboration

Different forms of collaboration can be achieved, strategic alliances (Harrigan 1987), joint ventures (Harrigan 1987), networks (Alter & Hage, 1993; Ebers, 1997; Ebers & Grandori, 1997; Ebers & Jarillo, 1997; Grandori & Soda, 1995; Jarillo, 1988; Miles & Snow, 1986; Snow, Miles & Coleman, 1992; Powell, 1987, 1991; Thorelli, 1986), cooperatives (Cornforth et al. 1988) and clusters (Porter 1998).

The joint venture (JV) is probably the most studied form of inter-firm relationship. This form of collaboration has been used by firms to achieve organizational goals, expand abroad and compete (Agarwal & Ramaswami, 1992; Contractor & Lorange, 1988; Geringer, 1988, 1991; Gill & Butler, 1996; Harrigan, 1985, 1987; Hennart, 1988; Kogut, 1988; Lorange & Probst, 1987; Nooteboom, 1999; Ohmae, 1993). According to Bartol et al. (1995), a joint venture (JV) can be defined as “an agreement between two or more organisations to jointly produce a product or service” (99).

Strategic alliances (SA) have become a very popular form of organizational strategy, as a means of accessing technologies. Sharing costs and risks in a SA allows firms access to scarce resources (Park 1996). Some authors consider that strategic alliances are “relatively enduring inter-firm co-operative arrangements, involving flows and linkages that utilize resources and/or governance structures from autonomous organizations, for the joint accomplishment of individual goals linked to the corporate mission of each sponsoring firm” (Parkhe 1991: 581). Lei and Slocum (1992: 81) identify alliances as “coalignments between two or more firms in which the partners hope to learn and acquire from each other the products, skills and knowledge that are not otherwise available to their competitors.” Myhrvold et al. (1995: 21) define a strategic alliance as “goal-oriented co-operation between two or more businesses involving a mutual exchange of resources and/or concerted efforts to resolve problems, based on formal agreements, coopting and minority stake investments—in contractual form.” Lorange and Roos (1997: 265) conceive a strategic alliance as “positioned between traditional wholly-owned organizations on the one hand and market-based interaction among several firms on the other.”

D’Cruz (1993) considers that in order to develop competitiveness, large organizations will establish networks with other organizations in order to transfer many of their operations to the network members while retaining strategic leadership of the network. Murakami (1993) understands a competitive network in terms of relations between a group of firms that co-operate to compete. Castells (1996) notes that in an era of new information technologies, a new organizational form has emerged, the ‘network enterprise’. Accordingly, a network enterprise is “that specific form of enterprise whose system of means is constituted by the intersection of segments of autonomous systems of goals” (Castells 1996: 171). He considers that “networks are the fundamental stuff of which new organizations are and will be made” (Castells 1996:168). As a simple primary definition, a network can be understood as a long-term relationship of a group of actors to achieve a common goal (Porras 2003).

What are clusters?

The idea of clusters is not new; early 20 century Marshall (1919) noted the importance of location for countries like France and Germany in comparison to the UK. During the decades of 1970s and 1980s in Italy emerged the idea of industrial districts “geographically defined productive systems (Pyke et al. 1992: 2).

However, it was late 1990s the idea of clusters became a fashionable topic due to the work of Porter. According to Porter (1998: 78), “clusters are geographic concentrations of interconnected companies and institutions in a particular field. Clusters encompass an array of linked industries and other entities important to competition.” In this perspective, geographic location is an important element in defining clusters. Porter considers that establishing clusters generates knowledge and could generate innovations for creating a competitive advantage, although some discontinuities can also emerge between participants due to the different levels of development, resources and capacities are different in each case.

Roelandt y Hertog (1999) consider that clusters focus on linkages and interdependencies between organizations in a value chain. Organizations in clusters have complementarities in horizontal or vertical activities, both needed for innovation. Mayer (2005: 41) understands industry clusters as a “group of firms that, a through their interactions with each other and with their customers and suppliers, develop innovative, cutting-edge products and processes that distinguish them in the market place from firms in the same industry found in other places”

Clusters allow SMEs to take advantage of the name or activities of larger organizations that otherwise could not make allowing them to have a better market image, to have access to better market opportunities and better contracts also; it allows them to take advantage of the synergies created by participants in clusters.

Clusters may include competitors, suppliers, government institutions, universities, research centers, trade associations, etc … This paper presents a case study of the automotive industry cluster located in San Luis Potosi, Mexico.

The State of San Luis Potosi

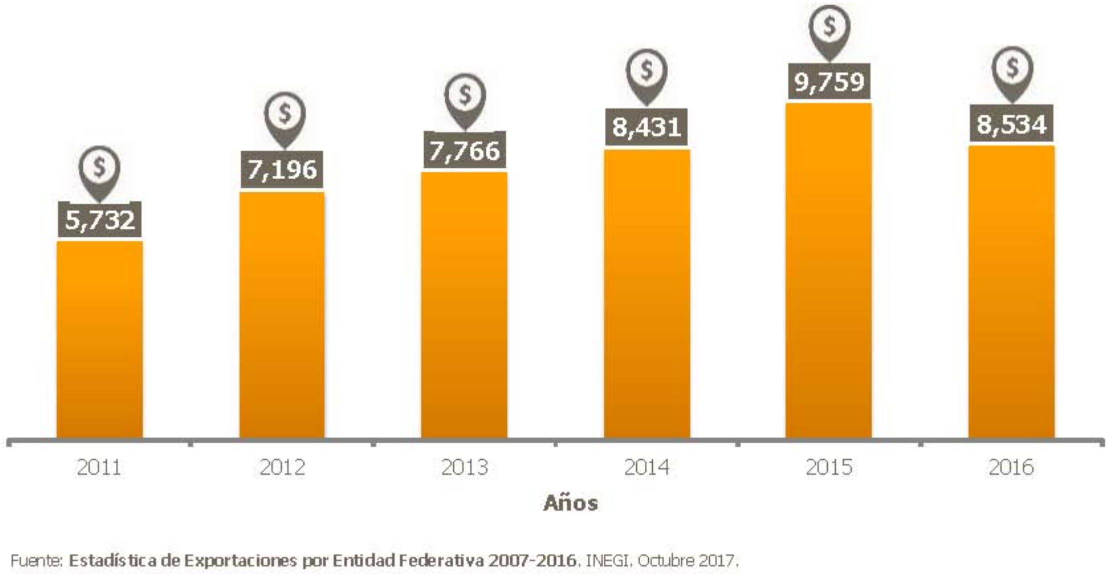
Located in the center East of Mexico, the State of San Luis Potosi together with the states of Aguascalientes, Guanajuato and Queretaro forms part of the region known as El Bajio, which, according to John Tutino (2016)[1], is one of the most dynamic regions worldwide. Some of the reasons include: best aeronautical cluster in Latin America; leading automotive exporter in Latin America, 4th worldwide, more that 10 assemblers and 800 suppliers; foreign direct investment from 80 countries; pharmaceutical leader in Latin America; more than 100 research centers, 200 universities and leader in industry 4.0; more than 146 specialized industrial parks, largest industrial corridor in Latin America; wine corridor, 76 vineyards.

In 2016 the World Bank ranked the city of San Luis Potosí as the eighth-best city to do business throughout the whole of Mexico; in 2015 KPMG nominated it as the fourth-best city for investment in the country. The area has become a hub for Mexico’s manufacturing industry, and in particular a center for automotive manufacturing. The state receives an important amount of FDI. Exports have witnessed consistent growth since early 2009. Where there are different industries located along the state like food and home appliances, the automotive industry has seen a boom with the arrival of various assembly plants including GM in 2006 and BMW in 2016 (OBG 2017).

The following graphics give an idea of the situation of the State of San Luis Potosi in different economic aspects including the evolution of exports and in particular manufacturing exports the period 2011-2016.

[1] Professor at School of Foreign Service Department of History, University og Georgetown.

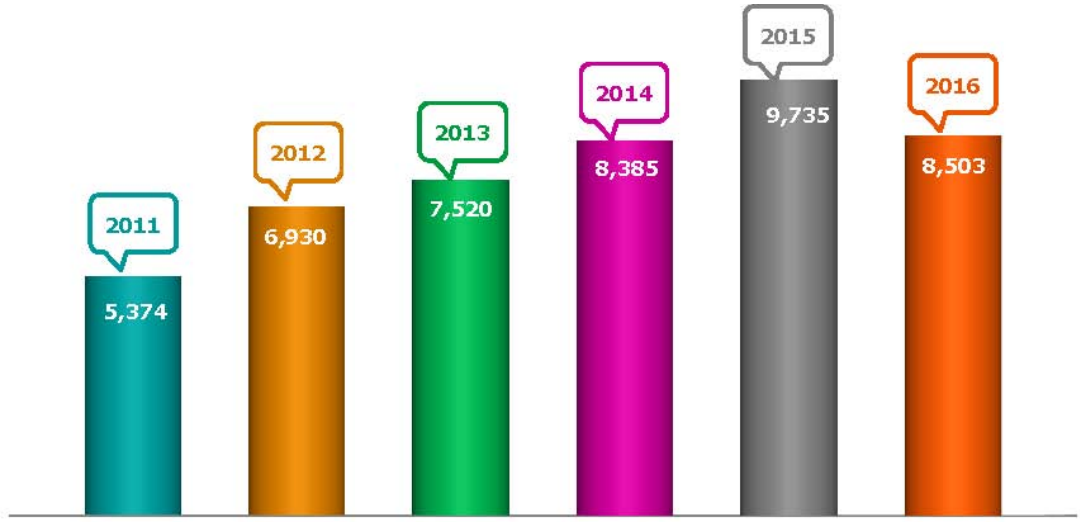
Graphic 1: San Luis Potosi total exports 2011-2016 (million dollars)



Graphic 2: San Luis Potosi growth annual rate exports 2011-2016

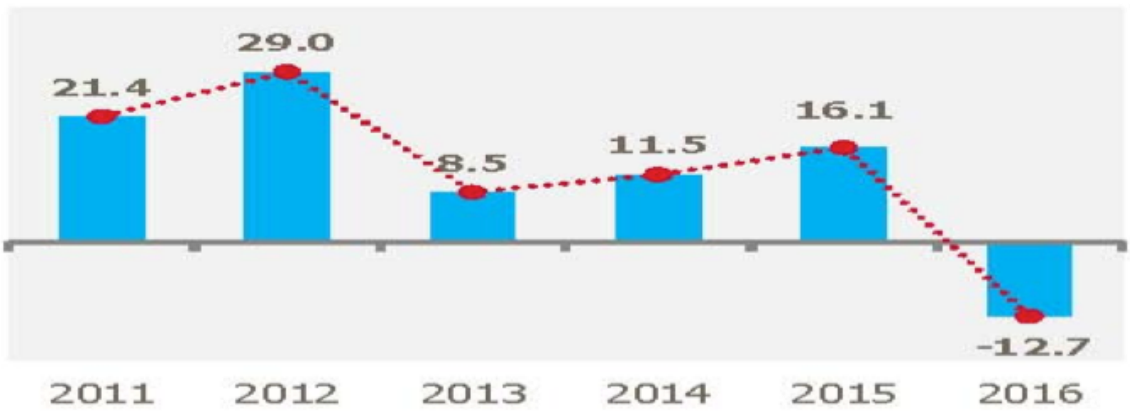
Source: Secretaría de Desarrollo Económico SLP based on INEGI October 2017

Graphic 3: San Luis Potosi manufacturing exports 2011-2016 (million dollars)



Fuente: Estadística de Exportaciones por Entidad Federativa 2007-2016. INEGI. Octubre 2017.

Graphic 4: manufacturing exports growth annual rate 2011-2016



Source: Secretaría de Desarrollo Económico SLP based on INEGI October 2017

According to official data, in the state of San Luis Potosi, there are 6 industrial parks registered, three clusters (automotive, logistics and medical), more than 229 automotive companies which represent more than 25% of manufacturing state GDP employing 82,000 persons.

The automotive industry cluster

The cluster was formed in 2015 in a triple helix scheme where public sector, industry and academia participate. Currently it is conformed in total by 35 enterprises including GM, BMW and 29 automotive companies, universities and research centers.

Some of the enterprises participating in the cluster include[2]:

- General Motors: Produces Chevrolet Trax, Equinox, Onix y GMC Terrain; transmissions; having 584 robots. GM reuses 100% of the manufacturing water employed in the plant contributing with the environment oriented towards the circular economy (Perez 2020).

[2] The information for this section was obtained from the official web pages of the different organizations and the cluster.

- BMW: announced the investment for facilities in the State in 2014. The plant was formally opened in June 2019 producing the BMW Series 3. In August 2020 produced the first 330e hybrid connectable. In August 2020 celebrated the production of the unit 50,000 produced in the plant that are exported to 36 countries. BMW uses Industry 4.0 in its production. According to official data, this is the most innovating plant of the group. In May 2021 the group officially announced the production of the Series 2 Coupé in San Luis Potosi starting summer 2021. 90% of the production process is automatized, using 500 robots. The plant makes use of the Smart Maintenance Assistant software assisting in the planning for equipment maintenance.
- Cummins motors: in 1980 establishes an agreement with DINA and in 1984 establishes its plant in San Luis Potosi. Currently the 95% of engines are for exporting.
- BrogWarner arriving in 1997: currently leader in propulsion systems for combustion vehicles, hybrids and electrics.
- Some other enterprises participating in the cluster include ABB, BOSH, GOODYEAR.

Cluster members share resources and experiences and for increasing knowledge and promoting innovation. The Industry 4.0 and Circular Economy are emerging as a result of organizations collaborating.

Industry 4.0

The 4th industrial revolution or Industry 4.0 surges in Hamburg, Germany in 2011 passing fro Information Systems / Information Technology to intelligent products, linked to: physic-cybernetic systems, the cloud, internet of things, big data, intelligent and dynamic production systems, high levels of communication, robotics, 3d printing. (Paiva & Charrua-Santos 2018)

Some principles for developing Industry 4.0 include interoperability, virtualization, decentralization, real time capacity, services oriented and modularity. Both, GM and BMW are making use of Industry 4.0 in its production processes. We can expect that other organizations participating in the cluster will follow their way in an institutionalization process (Meyer & Rowan 1977).

This new industrial paradigm is based on individual production, horizontal integration, collaborative networks and digital integration with supply chain (Klingenberg & do Vale Antunes 2017).

Circular Economy

According to NSW Circular (2020: 3), “Circular Economy is a systematic approach to economic development that benefits business, society and the environment… is regenerative by design and aims to gradually decouple growth from the consumption of finite resources. It is an accelerator to a zero-carbon sustainable economic recovery”. Circular economy moves organizations to a new world vision, zero waste, zero emissions, and transition to renewable energy.

As in the precious case, GM and BMW but not only them are addressing their production to a circular economy. As we mentioned, GM reuses 90% of the manufacturing process, BMW uses a painting section where processes do not generate residual waters at all. The water used in the process is treated and reused. Renewable energies will be used with a solar energy system located in the facilities.

Conclusions

Collaboration is a need for organizations to survive in a global, dynamic, turbulent environment. Different strategies can be used for achieving collaboration. In this paper we review clusters as a way organizations take for obtaining their goals. By collaborating in clusters organizations look for grow and innovation that can be achieved connecting with other organizations; technology transfer; access to human development, new skills and knowledge, training; innovation, sustainability and circular economy.

Clusters are a successful strategy for organizations, not only for SMEs but also for large organizations as we describe in this paper for the cases of GM and BMW among others in the State of San Luis Potosi. Clusters attracts Foreign Direct Investment to the region were there are located.

References

Agarwal, S. & Ramaswami, S. N. (1992). "Choice of foreign market entry mode: Impact of ownership, location and internalization factors". *Journal of International Business Studies*, Vol. 23(1): 1-27.

Alter, C. & Hage J. (1993). *Organizations Working Together*. Newbury Park: SAGE Publications.

Astley, W. G. (1984). "Toward an Appreciation of Collective Strategy". *Academy of Management Review*, Vol. 9(3): 526-535.

Astley, W. G. & Fombrun, C. J. (1983). "Collective Strategy: Social Economy of Organizational Environments". *Academy of Management Review*, Vol. 8(4): 576-587.

Axelrod, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.

Bartol, K, M., Martin, D. C. Tein M. & Matthews, G. (1995). *Management. A Pacific Rim Focus*. Sydney: McGraw-Hill Book Company.

Bresser, R. K. F. (1988). "Matching Collective and Competitive Strategies". *Strategic Management Journal*, Vol. 9: 375-385.

Bresser, R. K. F. & Harl, J. E. (1986). "Collective Strategy: Vice or Virtue?" *Academy of Management Review*, Vol. 11(2): 408-427.

Buttery, A. E. & Buttery, E. (1994). *Business Networks. Reaching new markets with low-cost strategies*. Melbourne: Longman Australia Pty Ltd.

Castells, M. (1996). *The Rise of the Network Society*. Great Britain: Blackwell Publishers Ltd.

Contractor, F. J. & Lorange, P. (Editors) (1988). *Cooperative Strategies in International Business*. Lenxington, Mass: Lenxington Books.

Cornforth, C. (1988). *Developing Successful Worker Co-operatives*. London: Sage.

D'Cruz, J. R. (1993). "Business networks for global competitiveness". *Business Quarterly*, Vol. 57(4): 93-98.

Ebers, M. (Editor) (1997). *The Formation of Inter-Organizational Networks*. Oxford: Oxford University Press.

Ebers, M. & Grandori, A (1997). "The Forms, Costs, and Development Dynamics of Inter-Organizational Networking" in Ebers, M. (Editor), *The Formation of Inter-Organizational Networks*. US: Oxford University Press.

Ebers, M. & Jarillo, J. C. (1997). "The construction, forms, and consequences of industrial networks". *International Studies of Management and Organization*, Vol. 27(4): 3-21.

Emery, F. E. & Trist, E. L. (1965). "The Causal Texture of Organizational Environments". *Human Relations*, Vol. 18: 21-32.

Gambetta, D. (1988). "Can We Trust?" in Gambetta, D. (Editor), *Trust. Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell.

Geringer, M. J. (1988). *Joint Venture Partner Selection: Strategies for Developed Countries*. Westport, Conn.: Quorum Books.

----- (1991). "Strategic Determinants of Partner Selection Criteria in International Joint Ventures". *Journal of International Business Studies*, Vol. First Quarter: 41-62.

Gill, J. & Butler, R. (1996). "Cycles of Trust and Distrust in Joint-ventures". *European Management Journal*, Vol. 14(1): 81-89.

Gray, B. (1989). *Collaborating. Finding Common Ground for Multiparty Problems*. San Francisco: Jossey-Bass Publishers.

Hamel, G., Doz Y. L. & Prahalad, C. K. (1989). "Collaborate with Your Competitors -and win". *Harvard Business Review*, (January-February): 133-139.

Harrigan, K. R. (1985). *Strategies for Joint Ventures*. Lenxington, Mass.: Lenxington Books.

----- (1987). "Strategic Alliances: Their New Role In Global Competition". *Columbia Journal of World Business*, Vol. 22(Summer): 67-69.

----- (1988). "Joint Ventures and Competitive Strategy". *Strategic Management Journal*, Vol. 9(2): 141-158.

Jarillo, J. C. (1988). "On Strategic Networks". *Strategic Management Journal*, Vol. 9: 31-41.

Klingenberg , C. O. & do Vale Antunes J. A. Jr. (2017). Industry 4.0: what makes it a revolution? Conference Paper · July

Kogut, B. (1988). "Joint Ventures: theoretical and empirical perspectives". *Strategic Management Journal*, Vol. 9(2): 319-332.

Larson, A. (1992). "Network Dyads in Entrepreneurial Settings: A Study of the Governance of Exchange Relationships". *Administrative Science Quarterly*, Vol. 37(March): 76-104.

Larson, A. & Starr, J. A. (1993). "A Network Model of Organization Formation". *Entrepreneurship Theory and Practice*, Vol. 17(2): 5-15.

Larsson, R. et al. (1998). "The Interorganizational Learning Dilemma: Collective Knowledge Development in Strategic Alliances". *Organization Science*, Vol. 9(3): 285-305.

Lei, D. & Slocum, J. W. Jr. (1992). "Global Strategy, Competence-Building and Strategic Alliances". *California Management Review*, (Fall): 81-97.

Lorange, P. & Probst, G. (1987). "Joint Ventures as Self-organizing Systems: A Key to Successful Joint Venture Design and Implementation". *Columbia Journal of World Business*, Vol. 22(Summer): 71-77.

Lorange, P. & Roos, J. (1997). *Strategic Alliances. Formation, Implementation and Evolution*. Great Britain: Blackwell Publishers Ltd.

Marshall, A. (1890). *Principles of Economics.* London: MacMillan and Co.

Mayer, H. (2005). “Cluster monitor”, *Economic Development Journal*, Fall pp. 40-53.

Meyer, J. W. & Rowan, B. (1977). "Institutionalized Organizations: Formal Structure as Myth and Ceremony". *American Journal of Sociology*, Vol. 83(2): 340-363.

Miles, R. E. & Snow, C. C. (1986). "Organizations: New Concepts for New Forms". *California Management Review*, Vol. XXVIII(3): 62-73.

Murakami, R. (1993). "Competitive networks: How to start and the benefits to expect". *Cost & Management*, Vol. 67(3): 36.

Myhrvold, T., Thomassen, G. P., Hofseth, R. & Langaker, P. O. (Editor) (1995). *Business network manual. A tool for developing business co-operation*. Oslo, Norway: The Norwegian Industrial and Regional Development Fund (SND).

Nooteboom, B. (1999). *Inter-Firm Alliances. Analysis and Design*. London: Routledge.

NSW Circular (2020). *The Circular Economy Opportunity*. Sydney. November.

Ohmae, K. (1989). "The Global Logic of Strategic Alliances". *Harvard Business Review*, Vol. 67(2): 143-154.

----- (1993). "The global logic of strategic alliances" in Bleeke, J. and Ernst, D. (Editors), *Collaborating to Compete. Using Strategic Alliances and Acquisitions in the Global Marketplace*. New York: John Wiley.

Oxford Business Group (OBG) (2017). The Report. San Luis Potosí 2017. www.oxfordbusinessgroup.com

Park, S. (1996). "The interfirm collaboration in global competition". Multinational Business Review, Vol. 4(1): 94-106

Parkhe, A. (1991). "Interfirm Diversity, Organizational Learning, and Longevity in Global Strategic Alliances". *Journal of International Business Studies*, Vol. Fourth Quarter: 579-601.

Paiva Santos, B., & Charrua-Santos F. (2018). *Industry 4.0: An Overwiew*. T.M. Lima Conference Paper · July.

Perez, R. S. (2020). “General Motors celebra 85 años de operaciones en México con más de 15 millones de vehículos producidos”, Jornada.

Polanyi, K. (1957). *The Great Transformation: The political and economic origins of our time*. Boston: Beacon.

Porras, S. (2003). “De la competencia a la colaboración: hacia nuevas formas de organización”, *Denarius* (8), pp. 99-124.

Porras, S., Clegg, S. & Crawford, J. (2004). “Trust as Networking Knowledge: Precedents from Australia” *Asia Pacific Journal of Management*, 21, 345–363.

Porter, M. Porter, Michael E. (1985). *Competitive Advantage*. New York: Free Press.

----- (1998). “Clusters and the New Economic of Competition”, *Harvard Business Review* (Nov-Dec), pp. 77-90.

Powell, W. W. (1987). "Hybrid Organizational Arrangements: New Form or Transitional Development?" *California Management Review*, Vol. 30(1): 67-87.

----- (1991). "Neither market nor hierarchy: network forms of organization" in Thompson, G. et al. (editors), *Markets, Hierarchies & Networks. The Coordination of Social Life*. London: Sage.

Pyke, F. (1994). *Small firms, technical servides and inter-firm cooperation*. Research series 99. Geneva, Switzerland: International Institute for Labour Studies.

Ring, P. S. & Van de Ven, A. H. (1992). "Structuring Cooperative Relationships Between Organizations". *Strategic Management Journal*, Vol. 13(7): 483-498.

----- (1994). "Developmental Processes of Cooperative Interorganizational Relationships". *Academy of Management Review*, Vol. 19(1): 90-118.

Roelandt, T. & Hertog, P. (1999). “Cluster analysis and cluster based policy making in OECD countries, an introduction to the theme”. In (OECD). Boosting Innovation The Cluster Approach, pp. 9-23.

Simon, H. A. (1947). *Administrative Behavior*. New York: Macmillan Publishing Co.

Snow, C. C., Miles, R. E. & Coleman. H. J. Jr. (1992). "Managing 21st Century Network Organizations". *Organizational Dynamics*, (Winter): 5-20.

Thorelli, H. B. (1986). "Networks: Between Markets and Hierarchies". *Strategic Management Journal*, Vol. 7: 37-51.

Tutino, J. (2016). *Creando un nuevo mundo. Los orígenes del capitalismo en el Bajío y la Norteamérica española*, México, Fondo de Cultura Económica, El Colegio de Michoacán, Universidad intercultural del Estado de Hidalgo, 2016, 832 pp. ISBN 978-607-164-209-7

Weber, E. P. & Khademian, A. M. (1997). "From agitation to collaboration: Clearing the air through negotiation". *Public Administration Review*, Vol. 55(5): 396-410.

Williamson, O. E. (1975). *Markets and hierarchies. Analysis and Antitrust Implications*. New York: The Free Press.

----- (1985). *The Economic Institutions of Capitalism*. New York: Free Press.

DECISION SCIENCES INSTITUTE

Unveiling Hidden Critical Suppliers in Deep Supply Networks

Francis Bowen	Janaina Siegler	Michael Schmierer	Sarah Harrison
Butler University	*Butler University*	*Conexus Indiana*	*Conexus Indiana*
Indianapolis, IN	Indianapolis, IN	Indianapolis, IN	Indianapolis, IN
fbowen@butler.edu	jsiegler@butler.edu	mschmierer@conexusindiana.com	sharrison@conexusindiana.com

ABSTRACT

This work expands on the application of the Nexus Supplier Index (Shao et al, 2018) in a dataset for 40 focal companies, representing 24 industries, 14 clusters, different sectors, sizes, generating $1.7T in revenue and 2.3M jobs. We collected ~16,000 datapoints from secondary data sources, covering upstream and downstream supply chains and calculated the NSI for the focal companies and their supply networks. We propose improvements for identifying hidden critical suppliers (HCS) that when combined provided visibility of depths up to 11-tiers. This data-driven approach deviates from traditional analysis in that it relies on the network topology, leveraging graph centralities.

KEYWORDS: Nexus Supplier Index, Hidden Critical Supplier, Supply Network, Supply Chain Analytics, Supply Chain Resiliency

INTRODUCTION

Supply chain resiliency is a trending topic due to the detrimental effect of disruptions on the economic wellbeing of every company linked to that network. In recent times, all industries have felt these disruptions on a global level, but little is understood about how to mitigate the dynamic flow of product and information through the complex supply networks. Take the recent semiconductor worldwide shortage as an example.

Since mid-December 2020, semiconductors have experienced a significant shortage. While the problem started in the traditional big name uses of semiconductors, this shortage has expanded to other industries including automotive and logistics companies (Pizzamento, 2021). There were also sanctions on a large manufacturing plant in China so many tried to stockpile before they were cut off from the supply. Texas manufacturer, NXP Semiconductors, was shut down in February 2021 due to weather and issues with their electric grid. Another large issue with this supply chain is the fact that many of the manufacturers that receive semiconductors work in a just-in-time supply chain; this creates little room for error, and they do not have inventory to spread around to others that may also need these supplies. Chip shortages cause widespread effects because they are in almost all consumer products, such as electronics, appliances, cars, medical devices, and technology in general. Along with decreased demand during the pandemic, there was also an increase in regulation, and major constraints put on the global transportation system.

Of the many factors affecting supply chain resiliency, visibility can have a profound effect on critical decisions for improving the readiness, response, and recovery from supply chain disruptions. Visibility of tier-1 suppliers gives only a narrow view of the entire picture. Recent literature provides evidence of the criticality of understanding suppliers in the deeper tiers that would otherwise go unnoticed but could have a strong impact on firms within the network.

Throughout this work, we may use the terms supply chain, supply network, supply web, value chain, supply universe in interchangeably ways. In any situation, we want to explicitly say that we consider the relationships among buyers and supplier in a complex, multi-tier, and multi-level intricated sets of relationships. We consider hidden critical suppliers those that are not necessarily visible to the focal companies, but they have the potential to disrupt the network and sometimes even entire industries.

With the underlying goal of cultivating supply chain resiliency through visibility, we chose to conduct a large-scale empirical study applying the Nexus Supplier Index. In the subsequent sections, the methodology, data collection methods, and a detailed look at the NSI calculation, will be given. Then a discussion of the results will summarize the key insights gleaned from the study, as well as our assessment of the NSI for identifying hidden critical suppliers in deep-tiered supply networks.

LITERATURE REVIEW

Supply chain resilience is a complex and multi-level construct. Some researchers highlight that to increase resilience, a company should focus on understanding and mitigating supply chain risks, which can have multiple definitions. According to Petit, Fikskel, & Croxton (2010), “Mathematically, vulnerability can be measured in terms of ‘risk’, a combination of the likelihood of an event and its potential severity.” Rao & Goldsby (2009) note that, “A situation is risky when it entails exposure to two essential components: exposure to an event and the uncertainty of possible outcomes”. The authors dive deeper into supply chain risk and break them out into five main factors and a total of 27 difference criteria.

At the same time that it is important that companies work on understanding and mitigating supply chain risks, It is also paramount that they increase their ability to fight those risks by increasing their own capabilities (Ponomarov and Holcomb, 2009; Pettit, Fiksel, and Croxton, 2010). Capabilities can be understood as, “Attributes that enable an enterprise to anticipate and overcome disruptions”. Capabilities could prevent an actual disruption (e.g., security measures deterring a terrorist attack), mitigate the effects of a disruption (e.g., stockpiles of emergency supplies), or enable adaptation following a disruption (Petit, Fikskel, & Croxton 2010).

Resilience is defined as, “The capacity for an enterprise to survive, adapt, and grow in the face of turbulent change” (Fiksel 2006). Petit, Fikskel, & Croxton further expand on the difference between risk management and resilience, which they believe is crucial for a company to grow and in turn build a competitive advantage against competitors. There is also another large group who understand that resilience should be understood by a foundational set of capabilities (Jüttner and Maklan, 2011) and that resilience develops in phases (Ponomarov and Holcomb, 2009). Further tying these concepts together to better understand and reduce risk to achieve a more resilient supply chain, Ponomarov & Holcomb (2009) suggest companies “Must be designed to incorporate event readiness, provide an efficient and effective response, and be capable of recovering to their original state or even better post the disruptive event (Ponomarov & Holcomb, 2009. p.)”. It is also important to know how to prioritize capability building. Jüttner and Maklan (2011) reiterate the research of Ponomarov and Holcomb (2009) that some capabilities can be seen as antecedents of supply chain resilience and called them formative capabilities. The formative capabilities are Flexibility, Velocity, Visibility, and Collaboration.

Flexibility is the firm's ability to pivot, velocity refers to how quickly a firm can respond and recover from an event, visibility is the knowledge a firm possesses about the rest of the supply network, and collaboration is the amount of mutual information shared with other firms.

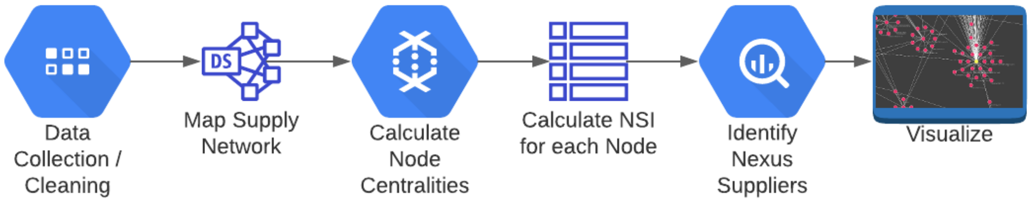
The resiliency of a firm can be evaluated based on their readiness prior to an event, their response to an event, and how they recover from an event. The four formative capabilities play a role in the three phases of resiliency, before, during, and after and event. According to Jutner et al (2011) velocity and all the capabilities that drive it, have the greatest impact on the three phases of an event that exposes a firm's resilience. It is also recognized that the formative capability groups are not mutually exclusive. Furthermore, the visibility of a firm supports a firm's velocity. It can be rationalized that the ability of a firm to move rapidly to prepare, react, and respond to an event, is a function of the information available of its own supply network. The criticality of visibility is central to other recent literature, where Glickman and White (2006) note security visibility is the crux of supply chain vulnerability mitigation, and Mandal (2017) provides an empirical study of visibility's dominant impact on supply chain resilience. Mandal's study offers insight from 207 supply chain professionals to further emphasize the significance of visibility and its role in resiliency. Moreover, the work by Mubarik et al. (2021), recommend improving resiliency through visibility by mapping the firm's supply chain.

Due to the aforementioned conclusions that visibility is a critical component for improving supply chain resiliency, we sought out methods for identifying hidden critical suppliers that could be reported back to the 40 selected focal companies. Yan et al. (2015) introduced the concept of a Nexus Supplier, which is defined as a supplier at any tier of a supply network that can have a profound effect on the entire network. In the traditional view of supplier relationships, the visibility of suppliers is limited to lower tiers and the impact from a firm's suppliers has a large effect on profitability. Conversely, the Nexus suppliers are those that have high potential impact on profitability and are not necessarily visible to the focal company. The work further defines three classifications of Nexus suppliers as operational, monopolistic, and industrial, each based on the position and structure of the supply network. Nexus suppliers are contingent on the overall network structure and adjacent relationships, regardless of tier. In the subsequent work by Shao et al. (2018), a data-driven approach is proposed to calculate a metric for identifying Nexus suppliers. The Nexus Supplier Index (NSI) is calculated for each firm in a supply network as an optimization problem, leveraging graph centralities for quantifying a firm's position in the network. The NSI then forms the basis for detecting Nexus suppliers.

THEORETICAL DEVELOPMENT/MODEL

The six-phase approach used in this empirical study is illustrated in Figure 1. All data collected for this work originated from public databases and was manually collected then cleaned for redundancy and standardization. Because data was aggregated from multiple sources, the same company may have different notations based on the source, requiring standardization. From the unified dataset, the supply networks could be mapped for a given focal company. Once the relationships are established for the supply network, key centralities were calculated for each node, which are the inputs for the NSI calculation. The NSI optimization is performed for each node in network, then finally the supply chain is visualized.

Figure 1. Summary of the methodology



Data Collection

Data was collected in the fall of 2020 from two public secondary data sources, Mergent Online and Bloomberg Terminals, with over 1,000 hours invested in the process, resulting in almost 16,000 data points. Starting from each of the 40 focal companies, data was collected in tiers. The focal companies were chosen to cover six main groups of industries that we called Conexus Group. The focal companies' real names were disguised to protect their privacy, but all the data is actual data and updated on May 2021. The companies are headquartered in 11 different countries, represent 24 industries, generate about $1.7 trillion dollars, 2.3 million jobs in 21K sites worldwide (Table 1).

Table 1: Focal companies data description

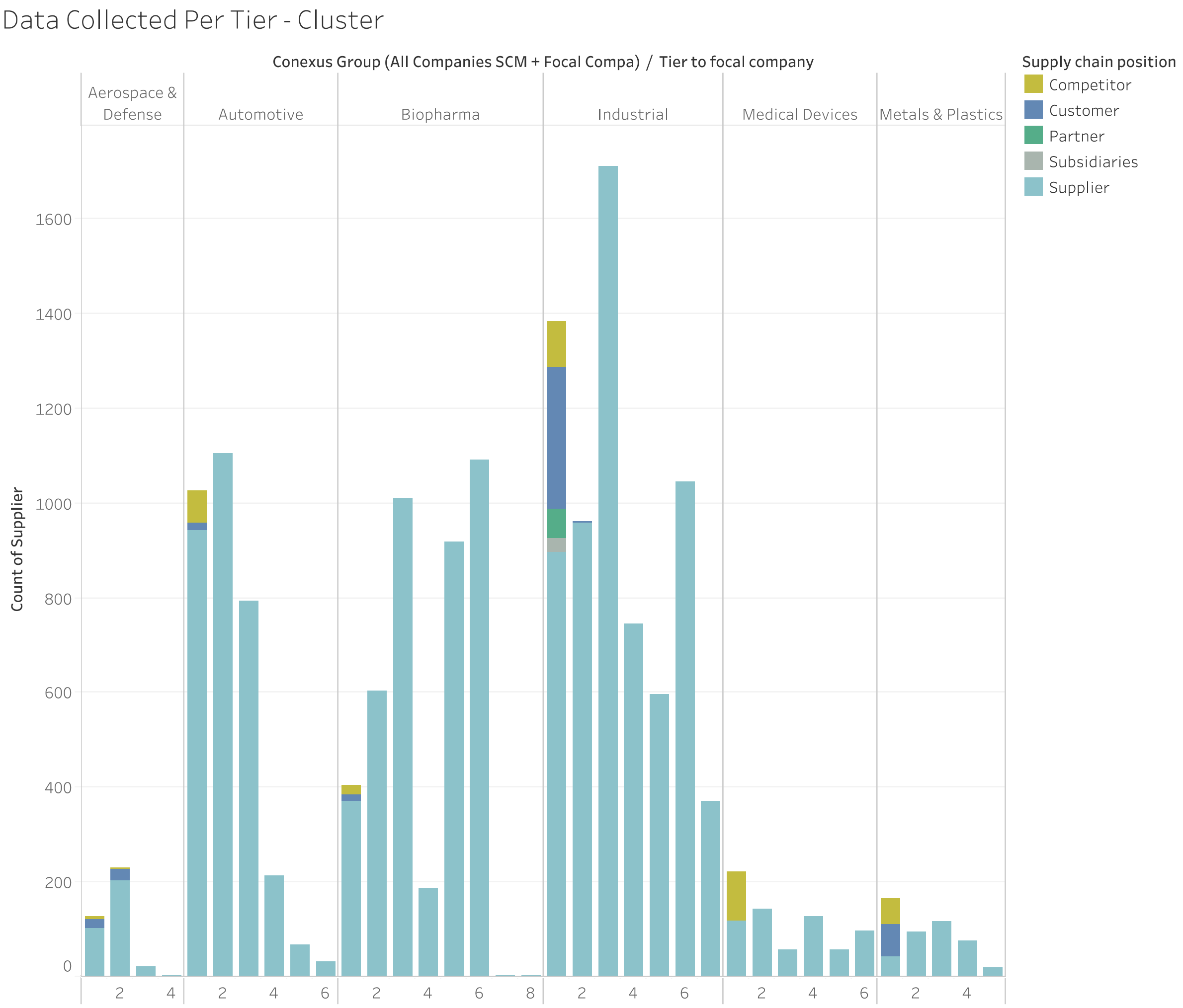
Conexus Group	HQ Country	Industry	Focal Company (Mask)	Corporation Annual Sales (2020)	Employees worldwide	Site worldwide
Aerospace & Defense	England	Aircraft engine and engine parts manufacturing	AD1	$21.00B	51.70K	358
	Ireland	Coating of metals and formed products	AD4	$28.00B	26.00K	3,329
	USA	Aircraft control systems, electronic	AD2	$1.50B	3.50K	
		Aircraft engine and engine parts manufacturing	AD3	$56.59B	181.00K	1,358
Automotive	Canada	Motor vehicle parts and accessories	AT3	$32.65B	158.00K	593
	Germany	Motor vehicle parts and accessories	AT9	$26.30B	68.00K	874
	Japan	Motor vehicle and car bodies	AT2	$30.00B	35.03K	588
			AT7	$137.55B	218.67K	1,579
			AT8	$256.00B	359.54K	2,604
		Motor vehicle parts and accessories	AT4	$34.86B	119.54K	266
	Netherlands	Motor vehicle and car bodies	AT5	$30.00B	77.00K	1,620
	USA	Motor vehicle and car bodies	AT6	$122.49B	155.00K	488
		Motor vehicle parts and accessories	AT1	$10.17B	49.70K	343
Biopharma	Switzerland	Pharmaceutical Preparation Manufacturing	BP5	$68.51B	101.47K	557
	USA	Chemical Laboratory apparatus, nec	BP1	$2.70B	10.00K	52
		Crop protecting services	BP3	$14.22B	21.00K	731
		Pharmaceutical Preparation Manufacturing	BP2	$3.27B	10.00K	85
			BP4	$24.54B	35.00K	278
Industrial	Japan	Industrial machinery and equipment	ID11	$19.00B	66.00K	613
	Switzerland	Measuring and controlling devices	ID7	$2.90B	14.40K	214
	USA	Chemical and allied products, nec	ID6	$0.41B	1.19K	14
		Construction machinery manufacturing	ID8	$41.75B	97.30K	1,037
		Gypsum Product Manufacturing	ID9	$2.49B	8.70K	143
		Internal Combustion Engines	ID4	$19.81B	57.83K	489
		Motor home manufacturing	ID10	$8.10B	17.50K	62
		Motor vehicle parts and accessories	ID1	$2.50B	7.70K	176
			ID2	$2.08B	3.70K	34
			ID3	$2.80B	10.50K	154
		Truck trailer manufacturing	ID5	$1.40B	6.75K	72
Medical Devices	China	Surgical and Medical Instrument Manufacturing	MD3	$0.23B	3.10K	27
	Denmark	Surgical and Medical Instrument Manufacturing	MD4	$0.53B	4.50K	25
	Japan	Fluid Power cylinder and actuator manufacturing	MD1	$4.80B	20.85K	316
	Usa	Burial caskets	MD6	$2.52B	11.00K	250
		Orthopedic, Prosthetic, and Surgical Appliances and Su..	MD5	$7.09B	19.90K	193
		Surgical and Medical Instrument Manufacturing	MD2	$2.68B	13.53K	77
Metals & Plastics	Ireland	Hardware, Not Elsewhere Classified	MP1	$2.72B	11.43K	30
	South Africa	Iron Ores	MP5	$11.00B	168.00K	1,065
	USA	All other plastics product manufacturing	MP2	$11.71B	47.00K	746
		Aluminum Rolling and Drawing, Nec	MP3	$9.29B	13.80K	109
		Steel Works, Blast Furnaces (Including Coke Ovens), an..	MP4	$10.40B	8.39K	122
Grand Total				**$1,066.54B**	**2,293.21K**	**21,671**

For each focal company, data was collected in all different directions and depth of tiers, considering downstream (customers), parallel (competitors), upstream (suppliers), as well as subsidiaries and partners. Table 2 summarizes the breakdown of the data collected per company in each direction and the depth we were able to go in terms of supply relationship. In one case, we were able to collect data up to Tier-8, in some cases we were not able to go beyond Tier-1, and on average we reach up to Tier-4 supplier information. Figure 2 provides the distribution of how much data was collected for the companies in each tier.

Table 2: Data collected per company

Conexus Group	Industry	Focal Company (Mask)	Max. Tier to focal company	Supplier (data points)	Subsidiaries/ Partners (data points)	Customers (data points)	Competitors (data points)	Total (data points)
Aerospace & Defense	Aircraft control systems, electronic	AD2	3	128		33		161
	Aircraft engine and engine parts manufacturing	AD1	1	3		3		6
		AD3	4	112		3	7	122
	Coating of metals and formed products	AD4	4	89		5		94
Automotive	Motor vehicle and car bodies	AT2	6	420			5	425
		AT5	4	557				557
		AT6	5	896				896
		AT7	6	278				278
		AT8	6	381				381
	Motor vehicle parts and accessories	AT1	4	152		8	6	166
		AT3	6	255		9	56	320
		AT4	5	106				106
		AT9	2	109				109
Biopharma	Chemical Laboratory apparatus, nec	BP1	6	2,480				2,480
	Crop protecting services	BP3	4	70		2	4	76
	Pharmaceutical Preparation Manufacturing	BP2	3	390		8	9	407
		BP4	4	757		4	6	767
		BP5	8	528				528
Industrial	Chemical and allied products, nec	ID6	1	5		5	21	31
	Construction machinery manufacturing	ID8	7	2,688	64	65		2,817
	Gypsum Product Manufacturing	ID9	7	845	30	10	4	889
	Industrial machinery and equipment	ID11			100	10		110
	Internal Combustion Engines	ID4	6	566		18		584
	Measuring and controlling devices	ID7	1	5		21	15	41
	Motor home manufacturing	ID10	7	244			6	250
	Motor vehicle parts and accessories	ID1	1	9		6	26	41
		ID2	6	630		169	22	821
		ID3	6	101		5	4	110
	Truck trailer manufacturing	ID5	7	1,236				1,236
Medical Devices	Burial caskets	MD6	4	170				170
	Fluid Power cylinder and actuator manufacturing	MD1	6	83			22	105
	Orthopedic, Prosthetic, and Surgical Appliances and Supplies	MD5	6	318				318
	Surgical and Medical Instrument Manufacturing	MD2	2	33			24	57
		MD3	1				34	34
		MD4	1				23	23
Metals & Plastics	All other plastics product manufacturing	MP2	4	214		20	17	251
	Aluminum Rolling and Drawing, Nec	MP3	5	66		3	17	86
	Hardware, Not Elsewhere Classified	MP1	5	16		8	7	31
	Iron Ores	MP5	5	36		16	10	62
	Steel Works, Blast Furnaces (Including Coke Ovens), and Rol..	MP4	5	20		19	5	44
Grand Total			**8**	**14,996**	**194**	**450**	**350**	**15,990**

Figure 2: Data collected per tier per company



Nexus Supplier Index

The Nexus Supplier Index (Shao et al, 2018) is based on the Nexus Supplier Theory (Yan et al, 2018) and requires viewing supply networks as a set of nodes in a directed graph. As such, the NSI is formulated as an optimization problem where the objective function is defined by key graph centralities.

Let $G = (S, E)$ be a graph consisting of a set of M nodes, $S = \{n_1, .., n_M\}$, and N edges, $E = \{e_1, .., e_N\}$representing a supply network. For a given firm in the network, the NSI_f, such that $f \in S$, is defined as,

$$\max_{\propto,\beta,\gamma,\sigma} \frac{\alpha D_f + \beta B_f + \gamma V_f}{\sigma F_f} \tag{1}$$

s.t.

$$\frac{\alpha D_i + \beta B_i + \gamma V_i}{\sigma F_i} \leq 1, \forall i \in S \tag{2}$$

$$\propto, \beta, \gamma, \sigma \geq 0 \tag{3}$$

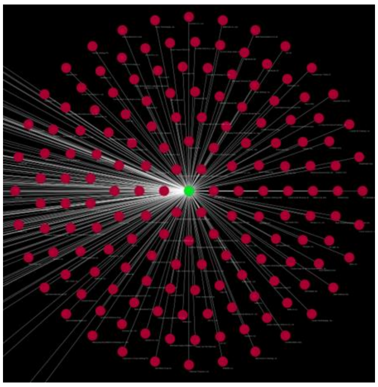
Where D_i, B_i, V_i, and F_i, are the degree, betweenness, eigenvector, and farness centralities, respectively.

The degree centrality for node i is defined as the number of links incident to a node in a graph and is defined formally in (4),

$$D_i = \sum_{j=1}^{M} A_{i,j} \tag{4}$$

Where $A_{i,j}$ is the adjacency matrix representation of graph G. In a supply network, a firm with a high degree centrality will have a strong influence on the overall network as a central hub for many other firms. Figure 3 provides an example node illustrating a firm with a high degree centrality.

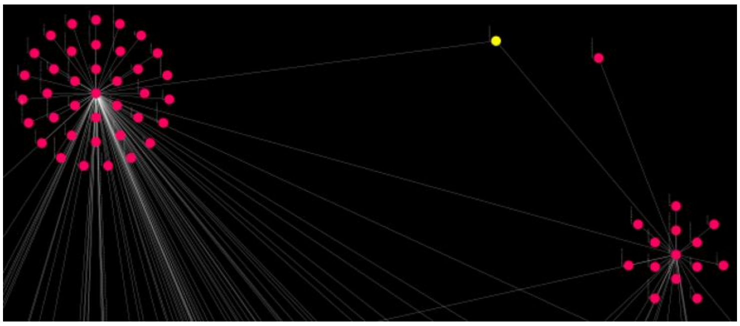
Figure 3. Firm in the Alcoa supply network with a high degree centrality



The betweenness central measures how frequently a node within a shortest path between two other nodes. Let P_s be the set of all shortest paths between any two nodes, and P_s^i be the set of shortest paths containing n_i, then the betweenness centrality for node i is defined as,

$$B_i = \sum \frac{|P_s^i|}{|P_s|} \tag{5}$$

Figure 4. Firm, highlighted in yellow, from the Alcoa supply network with a high betweenness centrality measure, linking two highly connected networks

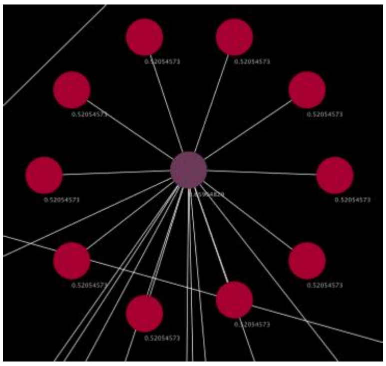


The eigenvector centrality measures the influence a node has within a graph, relative to the influence of its neighboring nodes. This centrality for n_i is the i-th component of the principal eigenvector, $x \in \mathbb{R}^M$, of the adjacency matrix, satisfying $Ax = \lambda x$.

$$V_i = \frac{1}{\lambda} \sum_{j=1}^{M} A_{i,j} \tag{6}$$

Although a high degree centrality can improve a node's eigenvector centrality score, if a node is connected to other nodes with low eigenvector scores, the overall influence of the node within the graph may be low and is exemplified in Figure 5. The principal eigenvector and corresponding spectral radius, λ, are solved in an iterative process.

Figure 5. Example firm in the Alcoa supply network with a low degree and high eigenvector centrality measures.



The closeness centrality measures the average shortest-path distance from a node to all other reachable nodes. The NSI calculation considers the reciprocal of closeness, farness, and is defined as,

$$F_i = \frac{\sum_{j=1}^{m-1} P_S^*(i,j)}{m-1} \tag{7}$$

Where m denotes the number of nodes reachable by n_i and $P_s^*(i,j)$ is the shortest path between n_i and n_j when n_j is reachable by n_i.

Shao et al. (2018) leveraged the NSI to define operational and monopolistic suppliers, where operational suppliers are regarded as having strong ties between the focal company and the supply network. In their study of the Honda supply network, the suppliers in the top 20% of degree, betweenness, and eigenvector centralities, for each tier, are viewed as operational suppliers. Moreover, they defined monopolistic suppliers as the suppliers in each tier that are in the top 10% in terms of the betweenness centrality measure. Monopolistic firms are critical to supply networks on the virtue that the companies operate on the greatest number of shortest paths and link together subnetworks.

RESULTS

Data was collected independently for each of the 40 focal companies from multiple public databases, in order of tier, and aggregated into a single dataset. For each company, direct links were collected, and percent revenue recorded when available. The breadth of available data varied significantly by industry, however once all data was aggregated, the size of the supply networks became extensive due to companies that operate in multiple industries. Figure 6 illustrates this phenomenon. Each of the 40 focal companies are listed along with number of data supplier relationships collected denoted in blue and the resulting size of their supply networks in orange. The maximum tier collected for any of the focal companies is 6, while the maximum tier discovered after aggregation is 12, with the average number of nodes per network was observed to be 4024.

The overall distribution of NSI scores for the entire dataset can be found in Figure 7, where the trends NSI score per tier is summarized in Figure 8. The overall mean NSI was found to be 0.548 with a standard deviation of 0.191. When observing the average NSI and standard deviation per tier, we observed a decreasing trend with increasing tiers.

Figure 6. The number of data points collected per focal company versus the resulting size of their supply network after aggregating all data.

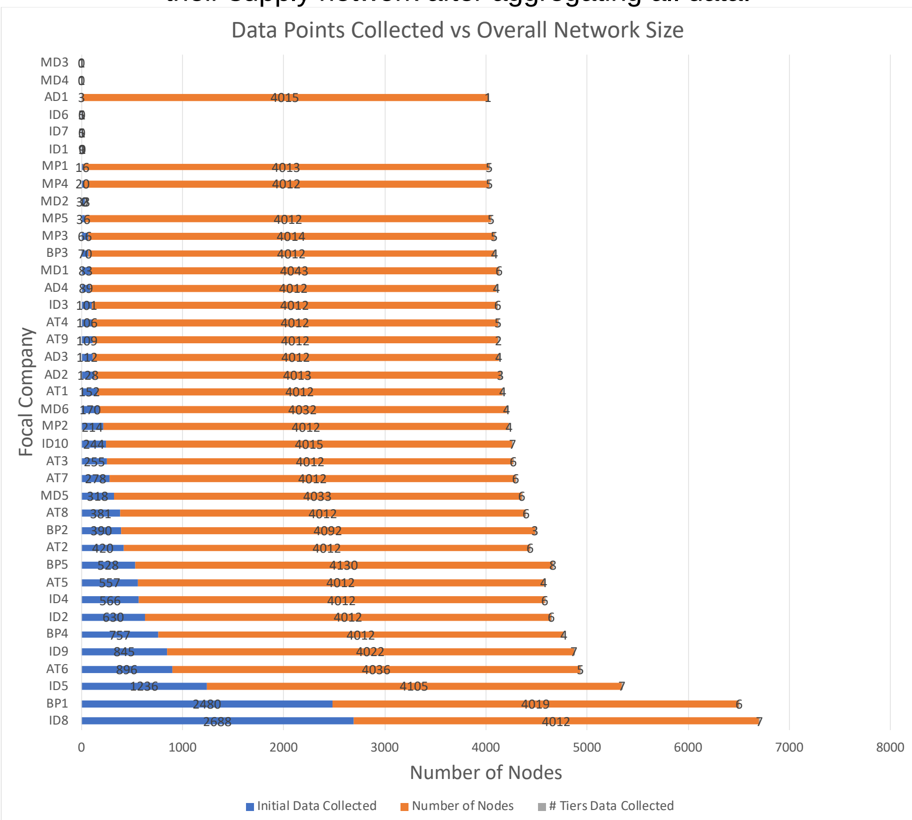


Figure 7. Distribution of NSI scores for the entire dataset

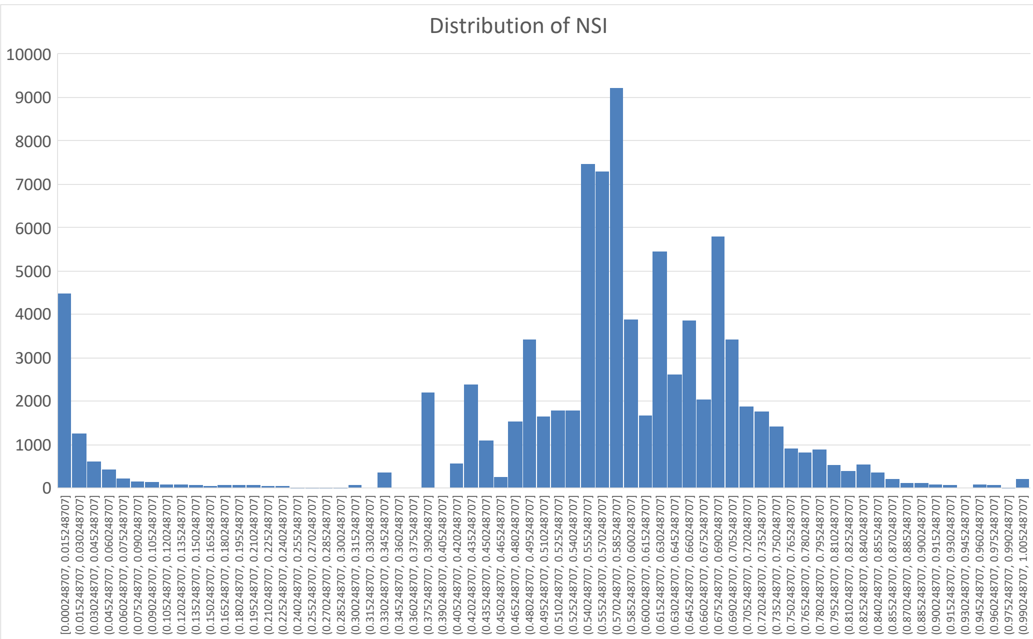
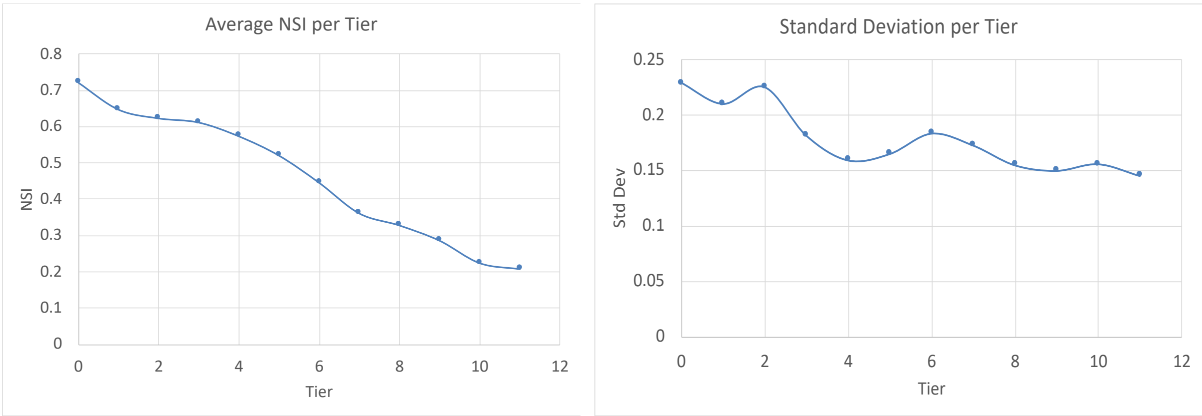


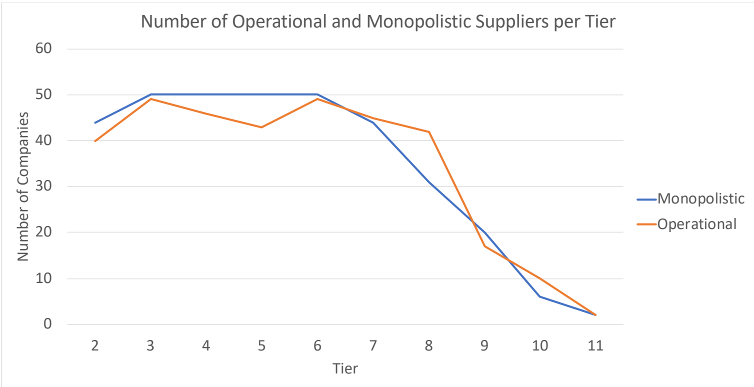
Figure 8. Average NSI per tier (a) and standard deviation per tier (b)



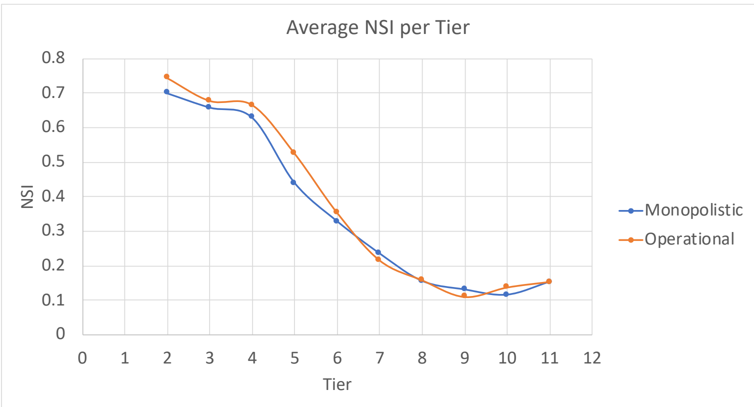
a. b.

Following the same criteria as the original authors, we identified operational and monopolistic Nexus Suppliers for each focal company then accumulated the results to observe the trends in NSI, which are summarized in Figure 9.

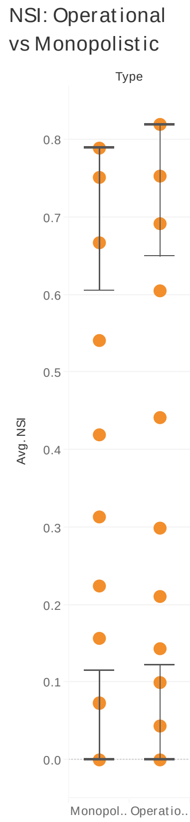
Figure 9. The number suppliers per tier (a), average NSI per tier (b), and the quartile summary (c), for operational and monopolistic Nexus Suppliers. The orange markers in the boxplot represent tiers with the highest tier at the bottom of the chart.



a.



b.

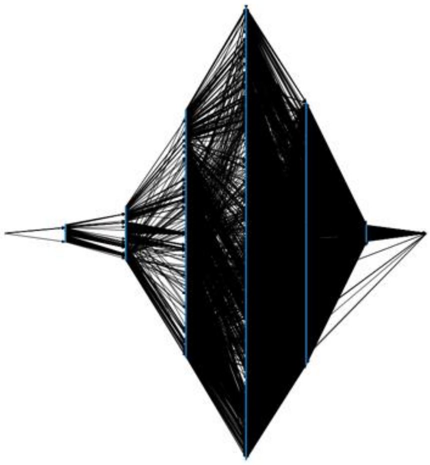


c.

DISCUSSION

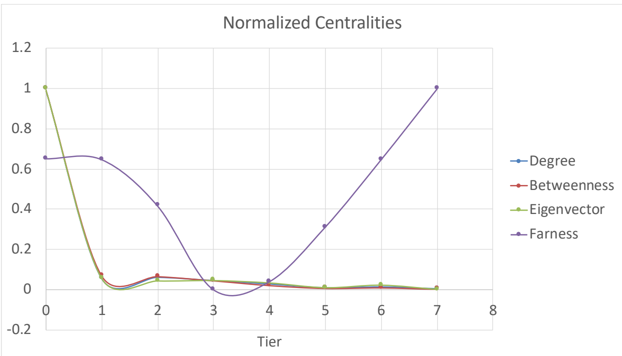
One of the first interesting findings was the shape of the overall network. Even though some authors have already started to map supply chains based on large raw datasets, this is still not the norm, specially going deeper in the supply chain. Similar to Kito et ao (2014), we were also able to find evidence that our supply network resembled a shape of a barrel. In other words, they had fewer nodes in the initial tiers, plenty in the intermediary tiers, and then fewer again on the more distant tiers. Our study, however, seems to be the first that involves multiple companies and multiple supply chains to test that concept. We sought out explanations to rationalize the observed trends in NSI per tier. As demonstrated in Figure 10, the supply networks for each focal company are found to have a barrel shape when viewed in order of tier.

Figure 10. Example supply network visualization showing a barrel shape when nodes are arranged by tier

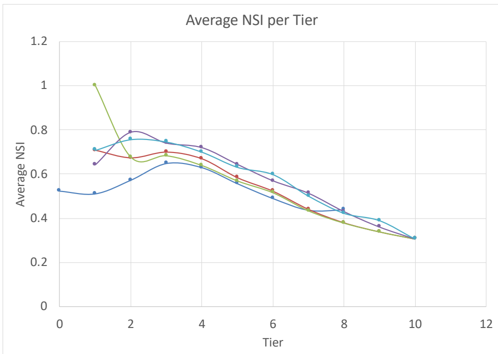


Additionally, we looked at the trends of the normalized centralities per tier for the identified Nexus Suppliers. For the example provided in Figure 11a, there is a small increase in degree, betweenness, and eigenvector measures for tiers 2 and 3, which are the largest tiers shown in Figure 10. This trend is reflected in the average NSI per tier for a randomly selected sample of Nexus Suppliers given in Figure 11b. The increase in degree, betweenness, and eigenvector centralities can be justified by the inherent barrel shape of the supply network, where the nodes in tiers 2 through 4 represent the firms with the greatest number of connections and are central to the entire supply network. Moreover, the farness measure is minimal in tiers 3 and 4 due to the higher concentration of nodes which contributes to a higher potential that a node resides on a shortest path to the remaining nodes. On either end point of the of the network, farness is greatest as those nodes have limited opportunity to contribute to a shortest path and reside the furthest from the opposite side of the network. With decreasing degree, betweenness, and eigenvector measures, coupled with a rapidly increasing farness measure, it is reasonable to conclude that the average NSI score will always decrease with increasing tiers.

Figure 11. Normalized centralities for the supply network in Figure 10 and a set of average NSI plots for five Nexus Suppliers



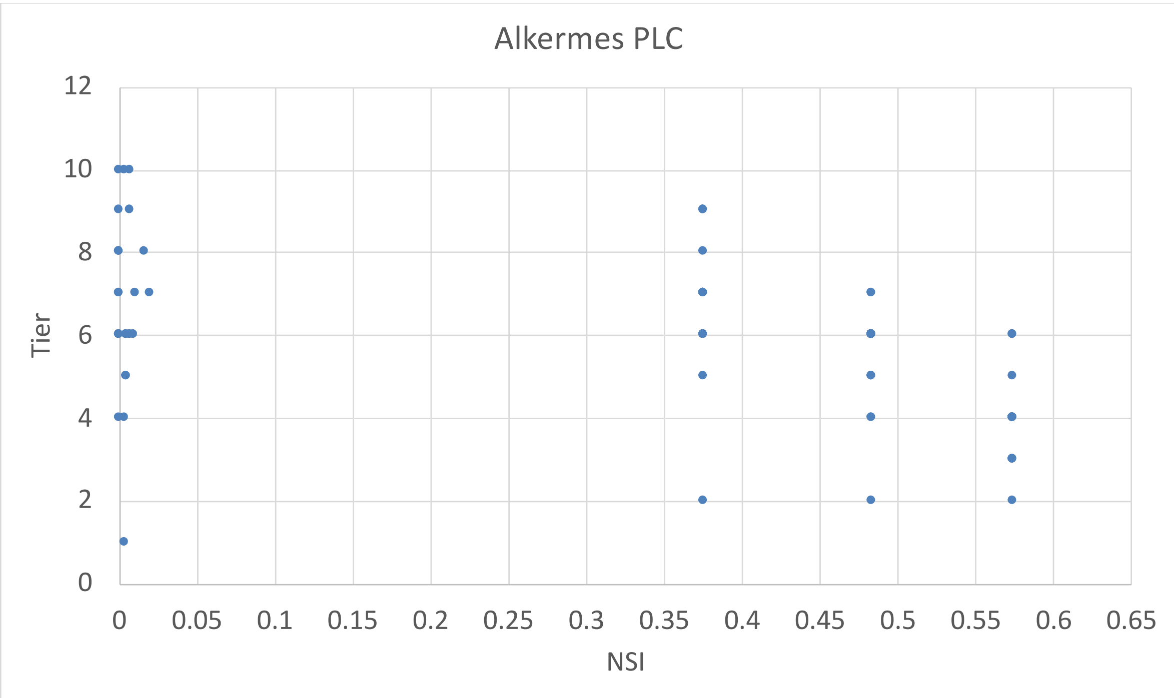
a.



b.

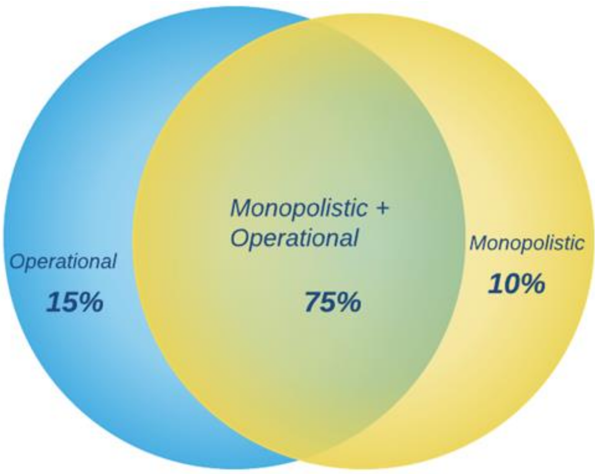
Due to the observed relationship between NSI score and tier, it is not practical to compare NSI scores across tiers which would explain why operational and monopolistic suppliers are identified on a per tier basis. Many firms appeared in multiple supply networks when data from all 40 focal companies was aggregated into a single dataset. Another key observation is that a company's NSI score can vary significantly based on different supply networks, which is expected because of the eigenvector centrality component of the NSI and variations in network topologies. One such example is Alkermes PLC, which appears in 23 of the 40 supply networks studied. Figure 12 plots the NSI score for Alkermes per tier, and illustrates the wide range of NSI scores, even along the same tier. For instance, Alkermes NSI score ranges from 0.015 and 0.57 when Alkermes appears as a tier 6 supplier. Factors that affect network topology include business cluster, industry, and sector types.

Figure 12. NSI score per tier for Alkermes PLC



Leveraging the NSI to identify operational and monopolistic suppliers is the primary motivation for exploring the approach. Visibility of the Nexus Suppliers provides focal companies the ability to understand key firms that have high potential for disrupting a supply network. Furthermore, these suppliers may not necessarily belong to the same industry or sector as the focal company, which is additional insight for how adjacent industries and sectors play a role in a focal company's operations. Figure 9a summarizes the number of operational and monopolistic suppliers per tier for the aggregated dataset, where we observed a significant number of suppliers that are classified as significant in the upper tiers of all the networks. The original criteria selects the companies in the top 20% of each centrality measure per tier for operational suppliers, and the companies in the top 10% of the betweenness measure for monopolistic suppliers. These definitions ensure a significant overlap between the two classifications and guarantees each type is identified on each tier. Figure 13 summarizes the amount of overlap of the two Nexus Supplier types with 75% of the Nexus Suppliers being identified as both operational and monopolistic.

Figure 13. Number of companies identified as operational and monopolistic



Both	408	75%
Monopolistic	52	10%
Operational	87	16%
Total	547	100%

The decreasing trend in the standard deviation of the NSI per tier, as summarized in Figure 8b, implies there should not be a static threshold for determining operational and monopolistic suppliers. There is negligible difference between the top and bottom centrality percentages for identifying the Nexus suppliers, therefore we propose identifying these suppliers as the outliers per tier. Regarding suppliers that have centrality measures that are at least one standard deviation above the mean as operational suppliers, and 1.5 standard deviations above betweenness as monopolistic suppliers, would resolve the issue of always identifying Nexus suppliers per tier. When the standard deviation is low, the NSI may not be a sufficient differentiating factor for identifying critical suppliers. Let O_t and M_t be the set of operational and monopolistic suppliers for tier-t, respectively. If N_t is the set of nodes with NSI scores two standard deviations above the mean for tier-t, then we define the set of Nexus suppliers for a supply network with T-tiers,

$$NS = \bigcup_{t=2}^{T}(O_t \cap M_t \cap N_t) \tag{8}$$

We recognize several other limitations with the work presented, in addition to how operational and monopolistic suppliers are identified. First, the results in this study are calculated on data from publicly available sources and each firm in the dataset is a publicly traded company. For completeness, direct collaboration with the focal companies would be necessary to validate the data and to include internal data, which will include private companies. Second, the NSI relies on shortest path calculations as key components to the betweenness and farness centralities, however there is not a guarantee that information or product can traverse that path within a supply network. Because publicly traded companies often own and operate subsidiaries, a separate analysis should be conducted on which firms in the supply network coupled through ownership or special partnerships. Lastly, the original NSI calculation strictly considers network topology. We propose a study in which the NSI calculation is performed by weighting the edges of the network with the relationships between the companies.

Contributions

In this work, we reinforced the power of the NSI for identifying hidden critical suppliers based on the topology of a supply network. The original authors applied this approach to Honda's supply chain to identify critical suppliers in tiers 3 and 4, whereas we applied the data-driven methodology to 40 companies representing a variety of industries with worldwide impact, 16,000 data points, with some supply networks as deep as 11 tiers. Moreover, our study spanned multiple countries with a diverse portfolio of focal companies. We have proposed an extension to the originally proposed NSI methodology for identifying suppliers which would otherwise be overlooked, and in doing so, we provide visibility to the deeper tiers of a supply network. While investigating the relationship between NSI and network tier, we generated evidence of the widely believed, but not proven, perception that supply chains are barrel shaped when viewed by tier. This was a significant finding because that network shape helps rationalize the trends in centralities and subsequent NSI score.

FUTURE WORK

In future work, we aim to extend the Nexus Supplier Index calculation by weighting the edges of the supply network based on the relationships of the firms, as well as exploring additional centrality measures. The data collection process will be automated to study the temporal trends, compare it to the NSI metric, and lastly, the quantitative results presented will be compared to a qualitative analysis. Ideally the secondary data can be combined with primary data from the focal company. Future collaborations will allow us to extend this methodology to specific focal companies that require insights into their hidden critical suppliers.

CONCLUSION

The Nexus Supplier Index is an intriguing metric for rating critical suppliers within a network. The score was selected as a means to identify hidden critical suppliers for 40 focal companies in the largest application of the NSI, reported in literature, with the intention of providing increased supply chain visibility to key worldwide companies. We studied the NSI trends over 24 industries and a myriad of different types of companies, ranging in size, in 11 different countries, that generate about $1.7 trillion dollars, 2.3 million jobs in 21K sites worldwide. The study demonstrated how the NSI score trends per tier and can be rationalized relative to the inherent barrel shape of each supply chain. We have proposed an extension of the NSI methodology for identifying operational and monopolistic suppliers, which includes a new proposal for the final set of Nexus Suppliers.

REFERENCES

Glickman, T. S., & White, S. C. (2006). Security, visibility and resilience: the keys to mitigating supply chain vulnerabilities. International Journal of Logistics Systems and Management, 2(2), 107-119.

Fiksel, J. 2006. “Sustainability and Resilience: Toward a Systems Approach.” Sustainability: Science, Practice and Policy 2(2):14–21.

Holton, G. (2004), “Defining risk”, Financial Analysts Journal, Vol. 60 No. 6, pp. 19-25.

Jüttner, U., Maklan, S. (2011) "Supply chain resilience in the global financial crisis: an empirical study", Supply Chain Management: An International Journal, Vol. 16 Issue: 4, pp. 246-259.

Kito, T., Brintrup, A., New, S., & Reed-Tsochas, F. (2014). The structure of the Toyota supply network: an empirical analysis. *Saïd Business School WP*, *3*.

Mandal, S. (2017). Supply chain resilience and internal integration: an empirical examination of different visibility categories. International Journal of Business Performance Management, 18(2), 216-235.

Mubarik, M. S., Naghavi, N., Mubarik, M., Kusi-Sarpong, S., Khan, S. A., Zaman, S. I., & Kazmi, S. H. A. (2021). Resilience and cleaner production in industry 4.0: Role of supply chain mapping and visibility. Journal of Cleaner Production, 292, 126058.

NXP. (n.d.). Our Company. https://www.nxp.com/company/our-company:OUR-COMPANY

Pettit, T.J., Croxton, K.L. and Fiksel, J. (2010), “Ensuring Supply Chain Resilience: Development of a Conceptual Framework,” Journal of Business Logistics 31, no. 1 (spring 2010): 1-22.

Pettit, T.J., Croxton, K.L. and Fiksel, J. (2013), “Ensuring Supply Chain Resilience: Development of a Conceptual Framework,” Journal of Business Logistics 34, no. 1 (March 2013): 46-76.

Pizzemento, A. 2021. The 2021 Semiconductor Chip Shortage: What, Why, and What's Next? Workforce Insights. Available at: < https://www.mau.com/workforce-insights/the-2021-semiconductor-chip-shortage-what-why-and-whats-next>

Ponomarov, S. and Holcomb, M. (2009), “Understanding the concept of supply chain resilience”, The International Journal of Logistics Management, Vol. 20 No. 1, pp. 124-143.

Shao, B. B., Shi, Z. M., Choi, T. Y., & Chae, S. (2018). A data-analytics approach to identifying hidden critical suppliers in supply networks: Development of nexus supplier index. *Decision Support Systems*, *114*, 37-48.

Rao S. and Goldsby, T. (2009), “Supply chain risks: a review and typology”, The International Journal of Logistics Management, Vol. 20 No. 1, pp. 97-123.

Ravi, S. (2020). CHIPS for America Act Would Strengthen U.S. Semiconductor Manufacturing, Innovation. Semiconductor Industry Association. https://www.semiconductors.org/chips-for-america-act-would-strengthen-u-s-semiconductor-manufacturing-innovation/.

Vakil, B., Linton, T., 2021. Why We’re in the Midst of a Global Semiconductor Shortage. Harvard Business Review. https://hbr.org/2021/02/why-were-in-the-midst-of-a-global-semiconductor-shortage

Yan, T., Choi, T. Y., Kim, Y., & Yang, Y. (2015). A theory of the nexus supplier: A critical supplier from a network perspective. *Journal of Supply Chain Management*, *51*(1), 52-66

DECISION SCIENCES INSTITUTE

Knowledge transfer from business practice to B-Schools:
A future research agenda based upon a case study of location technology adoption in supply chain curricula

William J. Ritchie, Ph.D.
James Madison University
Email: ritchiwj@jmu.edu

Joseph Kerski, Ph.D., GISP
Esri Inc. and University of Denver
Email: jkerski@esri.com

ABSTRACT

The purpose of this study was to explore knowledge transfer between practice and pedagogy in the supply chain content area of location analytics. Drawing upon a case study methodology, we explore spatial technologies in practice, in B-schools, and in supply chain curricula. We found that there is a significant chasm between location analysis in practice and B-school implementation. We provide theory and practical suggestions for B-schools and supply chain programs to catch up with practice. We also suggest a future research agenda to help address the chasm that exists between the practice of location analytics and B-school implementation.

KEYWORDS: Geographic Information System, Location Technology, Knowledge Transfer, Curriculum, Supply Chain, Analytics

INTRODUCTION

There is an extensive history of academic business research related to the transfer of knowledge from the academy to practice (Simmonds, Dawley, Ritchie, & Anthony, 2001; Steffens, Weeks, Davidsson, & Isaak, 2014) as well as within and between organizations (*Richter & Niewiem*, 2009; Ritchie, Drew, Srite, Andrews, & Carter, 2011). These literature streams have effectively examined technologies and theories as they are channeled from the academic environment through pedagogical methods and scholarly works to the practicing manager. However, there is limited attention in these research streams related to the transfer of knowledge from business practice back into the academic environment (Phan, Siegel, Wright, 2009). Anecdotal evidence suggests that knowledge transfer from practice back into the academic environment may occur through practitioner-grounded scholarly works that eventually find their way into mainstream higher education textbooks, classroom lectures, or similar resources.

However, given that the publication time horizons for peer-reviewed journals often require multiple years, classroom diffusion of knowledge derived from practice is significantly lagged (MacDonald & Richardson, 2011). Thus, the current pedagogical paradigm for coping with recent technological developments in practice is that Business School (or B-school) educators rely upon guest speakers, case studies, and experiential exercises to compensate for lagged or missing textbook content. However, textbooks remain the primary source of knowledge in the academic environment (Ferguson, et al., 2008), and like journals, textbook publishing is also a long process. Further, these incremental B-school classroom activities may carry some level of effectiveness when industry practices are evolving at a slow, incremental pace. However, there are frame-breaking shocks taking place in industry that challenge all of the current pedagogical paradigms. The most recent COVID-19 pandemic is one such shock that has driven the adoption of new technologies by entire municipalities, states, and countries. One of the technological centerpieces amidst the COVID-19 pandemic was location analytics; the digital mapping of the spread and impact of the virus globally. The recent proliferation of mapping pandemics, Big Data, Cloud Computing, and the Internet of Things (IoT) has received widespread industry attention. As such, it is a daunting task for B-school faculty to make sense of the applications, let alone identify a classroom activity to effectively teach these concepts to the students in a semester-by-semester or quarterly based academic environment.

The purpose of this study is threefold. First, we endeavor to identify a business case whereby a significant knowledge gap exists between practice and pedagogy. Second, building on the findings of Ramakrishna, Sarkar, and Vijayaraman (2011), we aim to provide B-school professors with tactics to aid in more rapid adoption and integration of nascent technologies into an existing curriculum. Finally, we suggest a future research agenda to help address the chasm that exists between the practice of location analytics in practice and B-school implementation. Overall, we seek to increase awareness among B-school faculties of the need to closely monitor technological developments in business practice to ensure that B-schools remain relevant to our students and the organizations that we serve.

Drawing upon a case-study methodology, we develop our model using data from recent developments in the adoption of location technologies in supply chain management. We endeavor to first chronicle the origins and development of location analytics technologies and the extent of adoption of supply chain spatial technologies in private industry and government. Next, we provide data related to current implementations of supply chain spatial technology in B-schools, followed by a more focused analysis of location technology content in popular textbooks. Finally, in our discussion we offer a theory for explaining barriers in the transfer of location analysis knowledge from practitioners back to B-school pedagogy. We also offer rapid adoption strategies for B-schools to “catch up” with practice.

LITERATURE REVIEW

Introduction to Location Technologies

To consider the presence of location technologies in practice, our case study centers on the recent diffusion and adoption of location analytic technologies in both private industry and government entities. Location analytics comprises a set of methodologies, tools, and data that enables the user to examine spatial patterns, uncover relationships, and solve problems related to changes over space and time. These components are part of an integrated set of technologies referred to as Geographic Information Systems (GIS), or in business, Location (spatial) Analytics. Through these technologies, data related to products, raw materials, or services related to a specific location can be examined in terms of flows, routing, and change by map scale or temporal scale. The data most commonly studied and produced through these analyses are in the form of maps, satellite and aerial imagery, infographics, tables, charts, and dashboards. The maps are increasingly web-based, interactive, and contain multimedia elements such as audio, video, text, and photographs, and are accessed via a cloud-based Software-as-a-Service (SaaS) model. The spatial data is increasingly accessed in a Data-as-a-Service manner, through open data portals and ArcGIS Hub technology. Location information is increasingly tied to IoT feeds, becoming instantly mappable and able to be analyzed through web-based GIS tools. This information grows in volume and variety, and a large percentage is delivered in real-time or near-real-time formats, such as the location of ships or vehicles, stream gauging stations, wildfire and other natural hazards locations and perimeters, traffic cams, and much more.

Location Technologies in Practice

A “location” component is also embedded in all fundamental business decisions and workflows, including adjusting supply chains on the fly, changes in routing of vehicles across a metropolitan area due to weather or construction, assessing risk from natural hazards, determining the optimal location for a new business or service, and many more. Efficient decision making due to location plays a significant role in the goal of achieving sustainability and resilience in business, government, academia, and nonprofit organizations.

As organizations increasingly are adopting location analytics tools and workflows, the variety of GIS software products in the market continues to expand. Esri’s ArcGIS platform holds the dominant market share, with recent embedding in the Microsoft PowerBI product, as well. However, other types of GIS software on the market include QGIS, Mapbox, and Leaflet. While these additional software platforms are gaining market share, our study of location technologies in practice centers on ArcGIS, due to the widespread adoption of Esri products worldwide. For example, of Fortune 500 companies, 50% use Esri GIS technologies. The adoption of GIS technologies is equally as extensive in the government sector, with most national governments and 20,000 cities, as well as all 50 states deploying Esri’s GIS/location technologies to evaluate and inform operations. In the educational sector alone, over 7,000 universities have adopted Esri’s location technology tools. In total, there are more than 350,000 organizations worldwide who rely on Esri spatial technologies (Esri, 2021). The most recent global COVID-19 pandemic is exemplar of such reliance on spatial technologies in practice. During the pandemic, various spatial representations, and maps of the spread of COVID were downloaded more than two trillion times from the Internet by mid-2021 and the majority of real-time COVID maps were hosted on Esri’s ArcGIS online platform.

METHODS

To address the research questions in this study, we employed a case study methodology. Case studies are effective for exploratory research and provide effective methods for theory building as well as multiple sources of rich qualitative data (Yin, 2003; Eisenhardt, 1989, Churchill, 1999, Alavi & Gallupe, 2017). We deemed the case study approach particularly appropriate for this study, since the scope of exploring knowledge transfer between practice and the academic environment requires multiple data points and a variety of perspectives. Case studies essentially provide greater opportunity for new insights that might otherwise be overlooked with single-source quantitative data approaches (Eisenhardt, 1989). Overall, our case study methodology, in conjunction with survey data, enables a deeper and personal examination of a phenomenon of interest within the environment where the phenomenon occurs. That is, the case-study approach afforded the opportunity to gain insights “from the field,” enabling the researchers to observe and study a phenomenon in real time.

The researchers identified four key criteria for the case study data collection used in this study to enhance the study’s construct validity (Gibbert, Ruigrok & Wicki, 2008). First, this study focuses on the examination of knowledge transfer within the broader field of supply chain management, due to the researchers’ intimate knowledge of the field and the technology being examined. Second, the authors focused on the adoption of location analytics within the supply chain industry. Importantly, spatial and location-based technologies are relatively new to the broader supply chain industry as well as to B-school curricula, providing a rich context for examining knowledge transfer.

Third, to accurately study knowledge transfer, our proxy for knowledge is “location” or “spatial” technologies. As such, we determined that the embodiment of location technologies must be sufficiently defined and recognized in both practice and the academic environments. The widespread adoption of spatial technologies, currently deployed in a majority of geology and cartography university departments, fulfilled this requirement. The clearly defined nature of the location analytics concept enabled an archival data analysis (Hill, 1993) of a sample of 12 mainstream classroom textbooks (see Appendix 1). Finally, the level of homogeneity in spatial technology adoption among organizations in the marketplace minimized the potential for spurious data that might arise due to differing interpretations of the term “spatial” or “location” technologies between firms. To guide this inquiry, the researchers followed the model presented by MacDonald and Richardson (2011) in their analysis of knowledge transfer from accounting practice to the academy.

Proxy for Location Analytics in B-Schools

With the increased adoption of location analytics and workflows in practice, there is an increased demand for employees with location analytics skills (World Economic Forum 2019 and 2021), yet despite the proliferation of GIS technologies and well documented business use cases in industry, a proportional level of adoption in B-schools is barely present. In fact, adoptions in higher education have taken a very circuitous route.

The use of GIS in higher education began in the late 1980s in university earth science and geography departments. In these cases, the location tools were used primarily for mapping the physical nature of the earth. By the mid-1990s, GIScience programs had emerged in many universities, and today, most universities and community colleges offer at least one course in GIS; others offer GIS certificates, and some offer bachelor’s and advanced degrees. Most of this instruction focuses on teaching *about* GIS, as a toolset and a science, known as geographic information science.

Beginning around 2010, with the advent of cloud-based GIS environments, open data portals, and web mapping tools, instructors in departments other than geography, environmental science, and GIS began to teach *with* GIS. This development began in the biological sciences, archaeology, and planning. These applications are spreading at a modest pace in health, civil engineering, digital humanities, and business (Ramakrishna et al. 2011). Business schools (Kerski 2020) are adopting location analytics, most notably in their supply chain management, information systems, marketing, and data science programs. However, in our review of B-school adoptions, the pace is very slow, and the implementation is spotty, primarily championed by a selected small group of faculty members within their respective colleges and universities. To obtain a more comprehensive understanding of the adoption and implementation of location technologies in B-schools, the researchers evaluated multiple information sources.
First, we conducted a general web search (using terms such as “location analytics”, “spatial analysis”, and “GIS”) as well as syllabi searches to identify B-school courses offering location analytics using GIS. Second, we evaluated the applied geography community (e.g. geography programs with some specialization in economics or business). Third, we examined generic organizational information from a GIS software provider, which included email inquiries and personal recommendations relating to B-schools deploying GIS technologies. fourth, the researchers’ applied their first-hand knowledge of B-school implementations by reviewing popular press and specialty GIS publications (Kerski, 2020). Finally, the researchers reviewed

12 mainstream supply chain textbooks that have been used in the higher education marketplace. The popularity of these texts was evidenced by the fact that all of the textbooks in our sample were in the 3rd edition or later. Three of the textbooks were in the 10th or later edition. A content analysis of the textbooks was conducted for terms such as “GIS software” and “spatial analysis software” in the subject indices and text.

RESULTS

Our analysis revealed 15 business schools in the United States and Canada that had integrated spatial analysis into their business programs. Only a few of these programs specifically mentioned using GIS or spatial analysis in a supply chain context. See Appendix 2 for a list of the schools. Our search also revealed only a few non-B-school college departments that taught business-related location analytics. These were not included in our sample, due to our focus on B-schools.

In our supply chain textbook content analysis, only a few of the textbooks mentioned location analysis software, and none of the texts provided solutions for common supply chain analysis problems (e.g. Vehicle Routing Problems (VRP), OD Matrix, Clustering, Service Areas, Distribution Networks, Supply Networks, etc.) using location technologies. For example, the primary tool for a common VRP was an Excel spreadsheet matrix.

DISCUSSION AND CONCLUSIONS

Our analysis revealed a significant chasm between location technology in practice and B-school implementations. After reviewing the case study data, we have found that location analysis is significantly lacking in B-schools, particularly in comparison to the widespread adoption in the marketplace. This corroborates the findings of Ramakrishna, Sarkar, and Vijayaraman (2011), who found minimal infusion of GIS technologies in B-schools nearly a decade prior to the current study. Viewed from Rogers (2003) diffusion of innovations theory, B-school adoption of location analytics is still primary led by what Rogers’ terms the “innovators” and “early adopters.” Rogers says that for an innovation to take root, the “early majority”, one standard deviation of adoptees below the mean, need to adopt it. Location analytics in business schools is still not used by this early majority.

As such, we offer the following possible research directions to help bridge this chasm. Our theoretical frame for this research is grounded in the adoption of innovations and bandwagon effects (Abrahamson & Rosenkopf, 1993), whereby technology adoption decisions are primarily driven by the actions of others. Specifically, we posit that *“…diffusion processes whereby organizations adopt an innovation, {are} not because of their individual assessments of the innovation's efficiencies or returns, but because of a bandwagon's pressure caused by the sheer number of organizations that have already adopted the innovation. (p. 488)*

Bandwagons require one of two frameworks to explain the adoption process: (1) fads, or (2) rational-efficiency perspectives (Abrahamson and Rosenkopf, 1993; Fiol and O'Connor, 2003). Fad theorists suggest that bandwagons are the result of either institutional (DiMaggio & Powell, 1983) or competitive (Abrahamson and Rosenkopf, 1990) pressures. Drawing upon this perspective, we might posit that the lack of adoption of spatial technologies is simply due to the fact that B-schools have not yet experienced pressure from either peer institutions or the marketplace to adopt such curricula. Conversely, we posit that the bandwagon pressures to retain current curricula are significant and prevent the inclusion of any new content. There is also an issue of timing, whereby the perceived need for the inclusion of spatial analysis might not have been viewed as necessary in the marketplace. However, the recent proliferation of spatial technologies, due to the COVID pandemic, is one such example of the timing component, whereby professors may now perceive the need for spatial technology adoption as a relevant addition to curricula, versus the past.

The rational-efficiency theorists focus on the economic efficiencies (Katz and Shapiro, 1985) whereby innovations are adopted due to superior information gathering capabilities, enabling more accurate assessments of the merits of a given innovation. Overall, researchers posit that bandwagons ensue due to other firm's beliefs that "hidden" economic benefits will be realized (Feder and O' Mara 1982; Poksinska et al. 2003; Viadiu et al. 2006). From this viewpoint we might suggest that B-schools do not have a robust business case for the adoption of the GIS or location technologies into the curricula and that the economic benefits have not been adequately assessed. For example, departmental budgets would be a consideration to determine if fiscal policies or resources preclude adoption of spatial technologies and GIS tools based upon cost/benefit analysis.

Overall, the bandwagon theory provides a basis for considering barriers to adoption. Thus, this stream of research might consider various aspects of inertia that might be present in curricula that would preclude the addition of new technologies. Similarly, those schools with active spatial technology programs might consider how to bolster adoptions to reach a bandwagon effect. Marketplace examples of such initiatives abound when considering the adoption of other administrative innovations such as quality standards (Ritchie & Melnyk, 2012).

An important complement to the discussion of whether to adopt spatial technologies is the perspective of undergraduate students. Thus, the researchers surveyed 159 undergraduate students who were enrolled in an entry-level supply chain course during a three-semester period from Spring 2020 through Spring 2021. Students were surveyed during a one-week delivery of an introductory session covering spatial analytics in supply chain (see Appendix 3 for a detailed summary of survey questions and responses). Students reported a strong interest in learning more about spatial technologies (GIS) during the introductory lesson with more than 80% of the survey respondents indicating that they "…wish they knew more about location analytics…" and that "…GIS tools are useful to business…" Overall, survey results indicated a very high level of interest in additional location analytic applications. Importantly, students also reported a very low level of awareness related to the application of location analytic tools to supply chain issues in practice. These preliminary results offer evidence for the need for future research related to student interest in the application of spatial analytic tools in supply chain applications and B-schools in general.

Having personally implemented spatial analytics into a current introductory supply chain course, the researchers offer the following recommendations for B-school professors and administrators. First, it is important to recognize that spatial analytics is not just for "marketing" or the "technology classes." As has been observed with the proliferation of performance dashboards during the COVID pandemic, spatial analytics touch many business disciplines. The more obvious supply chain applications such as vehicle routing and clustering applications, are equally impactful tools when applied to financial market concentrations or human resource concentrations geographically. Also important are the most recent applications of spatial analytics to cybersecurity (e.g. gas pipeline disruptions) and population demographics. Second, with all these integrated and overlapping applications, B-schools can readily make a case for applying GIS adoption across analytics, supply chain, entrepreneurship, marketing, management, and risk management. These applications are certainly not exhaustive, as virtually every facet of business education has a location component. Consider the growing use of Google and other smartphone technologies.

Finally, there are some simple, yet significant steps, that B-schools can take to begin the journey into spatial analytics. For starters, B-schools can offer basic incentives for professors to explore GIS applications. The spatial analysis platform discussed in this study has a well-developed cloud-based platform. As such, the GIS technology can be readily accessed in the classroom, regardless of existing university IT infrastructure or location. Furthermore, tools such as Business Analyst Web, that instructors and students can use online, are already a part of the Esri technology suite that is already on most university campuses. There are also many online resources on platforms such as Udemy, YouTube, and Esri.com that provide case studies in article and video form, lessons, and spatial data about spatial applications in a variety of fields, including supply chain.

The lessons provide step-by-step workflows that instructors could plug into their courses so that students work in a hands-on manner on the very activities that their instructors have just provided foundation on in their courses. For example, a hands-on lesson on risk assessment from Florida hurricanes could supplement instruction on how insurance rates are set and revised. A hands-on lesson on siting the optimal location for a store in a chain in a city could supplement instruction on the influence that zoning, competition, and demographic characteristics have on the location. In these hands-on activities, the students use the same type of data, including satellite imagery, traffic volume on streets, location of competitors, and consumer behavior, that people in the workplace use on a daily basis. They create derivative data and information such as drive-time rings or buffers to their proposed store locations, or infographics on demographic characteristics for identified neighborhoods. They present their findings graphically and in oral or video form using interactive web mapping applications such as story maps. These maps and data layers are shareable, so they can get feedback from their peers and from their professor, refining their analysis as time progresses, again, simulating the process that occurs in the workplace.

A growing number of location analysis textbooks and self-guided instructional manuals are emerging. In terms of financial resources, in one university in our sample, the geography department offered mini-grants to professors who were interested in applying GIS technologies in other academic disciplines. There are also opportunities to generate a special interest group in a Professional Society such as DSI for networking faculty together on spatial analytic analysis topics. Creation of a consortia of universities to research on the effectiveness and implementation of location analytics in B-schools is another possible avenue for development. This would accomplish for business what the National Center for Geographic Information and Analysis (NCGIA) did for geotechnologies in the early 1990s. The NCGIA grant-funded center comprising three universities turned GISystems into GIScience—from a set of tools and a scattering of instructors and campuses into a science, with its own community of practice, research agenda, and development goals. A location analysis education consortia of universities could hasten adoption, research, and development of these tools, workflows, and perspectives that are critically needed in academia and the workplace.

APPENDIX 1
Supply Chain Textbooks Reviewed

Supply Chain Management	4th Edition	Lambert
Supply Chain Management	6th Edition	Chopra & Meindl
The Supply Chain Game Changers		Burnette, Dittman, & Stank
Purchasing and Supply Management	15th Edition	Johnson & Flynn
Operations Management Processess and Supply Chains	11th Edition	Krajewski, Malhotra, & Ritzman
Operations Management: A Supply Chain Process Approach		Wisner
Principles of Operations Management: Sustainability and Supply Chain Management	10th Edition	Heizer, Render, & Munson
Supply Chain Logistics Management	4th Edition	Bowersox, Closs, Cooper, & Bowersox
Supply Chain and Logistics Management Made Easy		Myerson
Introduction to Operations and Supply Chain Management	4th Edition	Bozarth & Handfield
Supply Chain Management: A Learning Perspective	3rd Edition	Kim
Operations and Supply Chain Management	4th Edition	Jacobs & Chase

Appendix 2
B-Schools Implementing Spatial Technologies

West:
University of California San Diego
California Baptist University
University of California Riverside
University of Redlands
Colorado Mesa University
Arizona State University

Midwest:
Carroll University
Texas Christian University

East:
James Madison University
West Chester University
University of the Potomac
Georgia Tech University
Western Connecticut State University

Canada:
Ryerson University, ON
Mt Royal University, AB

Appendix 3
Supply Chain Course Student Comments

Question	Somewhat Agree	Agree	Strongly Agree	Total
I wish I knew more about location analytics	20%	39%	28%	81%
Business can greatly benefit from GIS tools	11%	31%	47%	89%
I have always wanted to learn more about "location analysis"	26%	24%	18%	68%
Supply Chain Managers should have a full knowledge of GIS technologies	14%	32%	38%	84%
I expect that I will use GIS technologies in my future employment	25%	27%	18%	70%
My future employers will be impressed to learn that I know GIS	22%	36%	24%	82%
GIS tools are useful to business	19%	36%	32%	87%
I intend to use GIS technologies in my career	21%	24%	18%	63%
I wish the college of business had more GIS-related courses	23%	19%	19%	61%
I think there are a lot of possible applications for GIS in business	22%	42%	22%	86%
GIS is useful in analyzing Big Data	19%	31%	18%	68%
Prior to enrolling in this class, I never thought About using GIS in business	28%	18%	13%	59%

Question	Strongly Disagree	Disagree	Somewhat Disagree	
I can readily explain the concept of location analytics to someone	16%	29%	12%	57%
I have heard other business students talk about how useful GIS tools are	14%	27%	13%	54%
I am familiar with the "VRP" solutions in the GIS software	27%	39%	10%	76%
I am familiar with the "service area" function in GIS software	28%	35%	10%	73%

REFERENCES

Abrahamson, E., & Rosenkopf, L. (1993). Institutional and competitive bandwagons: Using mathematical modeling as a tool to explore innovation diffusion. *Academy of Management Review*, 18(3).

Alavi, M, & Gallupe, R. B. (2017). Using information technology in learning: Case studies in business and management education programs. *Academy of Management Learning & Education,* 2(2).

Churchill, G.A.J. (1999). *Marketing research: Methodological foundations*. Fort Worth: The Dryden Press: Harcourt Brace College Publishers.

DiMaggio, P.J., Powell, W.W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2).

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.

Esri. (2021). *Media fact sheet*. https://www.esri.com/en-us/about/media-relations/fact-sheet

Gibbert, M., Ruigrok, W. & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, 29(13), 1645-1474.

Hill, M.R. (1993). *Archival strategies and techniques*. London: Sage Publications.

Katz, M.L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75,424–40.

Kerski, J. J. (2020). 5 Shining examples of the use of GIS in university schools of business. *Esri Industry Blog.* Esri: Redlands CA. https://www.esri.com/en-us/industries/blog/articles/gis-schools-of-business/

MacDonald, L.D. & Richardson, A.J. (2011). Does academic management accounting lag practice? A cliometric study. *Accounting History*, 16(4), 365-388.

Mills, C.W. (1959). *The sociological imagination*. Oxford: Oxford University Press.

Phan, P.H., Siegel, D.S., & Wright, M. (2009). New developments in technology management: Background issues, program initiatives, and a research agenda. *Academy of Management Learning & Education*, 8(3), 324-336.

Pick, J. (2008). *Geo-Business: GIS in the digital organization.* Wiley.

Ramakrishna, H., Sarkar, A. & Vijayaraman, B. (2011). A comparison of infusion of GIS and spatial analysis concepts among AACSB and Non-AACSB business schools. *Proceedings, Western Decision Sciences Institute.*
http://wdsinet.org/Annual_Meetings/2011_Proceedings/papers/Paper260.pdf

Richter, A., Niewiem, S. (2009). Knowledge transfer across permeable boundaries: An empirical study of clients' decisions to involve management consultants. *Scandinavian Journal of Management*, 25(3), 275-288.

Ritchie, W.J., Drew, S.A., Srite, M., Andrews, P., & Carter, J. (2011). Application of a learning management system for knowledge management: Adoption and cross-cultural factors. *Knowledge and Process Management*, 18(2), 75-84.

Rogers, E. (2003). Diffusion of Innovations, 5th Edition. Free Press, 576p.

Simmonds, P.G., Dawley, D.D., Ritchie, W.J., & Anthony, W.P. (2001). An exploratory examination of the knowledge transfer of strategic management concepts from the academic environment to practicing managers. *Journal of Managerial Issues*, 13(3), 360-375.

Steffens P.R., Weeks, C.S., Davidsson P. & Isaak L. (2014). Shouting from the ivory tower: A marketing approach to improve communication of academic research to entrepreneurs. *Entrepreneurship Theory and Practice*.

Viadiu, F.M., Fa, M.C., & Saizarbitoria, I.H. (2006). ISO 9000 and ISO 14000 standards: An international diffusion model. *International Journal of Operations and Production Management* 26(2), 141–65.

World Economic Forum. (2021). *Data excellence: Transforming manufacturing and supply systems.* http://www3.weforum.org/docs/WEF_Data_Excellence_Transforming_manufacturing_2021.pdf.

World Economic Forum. (2019). *Data science in the new economy: A new race for talent in the fourth industrial revolution.* http://www3.weforum.org/docs/WEF_Data_Science_In_the_New_Economy.pdf

Yin, R.K. (2003). *Case study research: Design and methods*, 3rd ed., Thousand Oaks: Sage.

DECISION SCIENCES INSTITUTE

Geospatial and Socioeconomic Analysis of Social Media Access and Purposeful Use in US Counties

Avijit Sarkar
University of Redlands
Email: avijit_sarkar@redlands.edu

James B. Pick
University of Redlands
Email: james_pick@redlands.edu

Tridev Raut
University of Redlands
Email: tridev_raut@redlands.edu

ABSTRACT

Purposeful use of the internet and particularly the expansive use of social media is rapidly shaping digital societies worldwide. This study examines spatial patterns and socioeconomic determinants of social media access and purposeful use for communication, commerce, entertainment, and influencing in U.S. counties. County-level analysis provides nuanced insights about the urban-rural social media digital divide and points to the importance of higher educational institutions as catalysts for social media usage. Age structure, urbanization, race/ethnicity and professional, scientific, technical services and overall service sector occupations are found to influence social network access and purposeful use. Implications of these findings are discussed.

KEYWORDS: Social media, Purposeful use, US county, Spatial analysis, and Regression

INTRODUCTION

Social media and social networks have become an integral component of the internet environment and overall digital society worldwide. Individuals, businesses, organizations including governments use social media to interact with peers and with each other. Social networks are used by individuals to connect with friends and family and social media is often leveraged as a form of self-expression. Local communities and governments often leverage social media for local contextual knowledge and expertise. Content created on social media can contain incredibly valuable information assets for a wide variety of stakeholders including individuals.

As social media platforms have proliferated, the population of social media users in the United States has continued to multiply and forms of social media use have also diversified. In the United States, the number of people 15 years or older who have used the internet to access and use online social networks has steadily increased from 137 million (70.6% of adults, 15+ using the internet) in 2015 to almost 160 million (73.7%) in 2019 (NTIA, 2019). Recent surveys have shown that social media interaction is part of the daily routine of 6 out of 10 Facebook users and 7 out of 10 Instagram and Snapchat users, who visit these sites at least once every day (Pew, 2021). In terms of popularity, YouTube and Facebook continue to dominate the social media landscape, followed by Instagram, Pinterest, LinkedIn, Snapchat, and Twitter (Auxier and Anderson, 2021). Interestingly though, with the exception of YouTube and Reddit, most social media platforms have showed little growth since 2019 until 2021 – a period that includes the COVID-19 pandemic. Surveys have also shown that discrepancies in social media access stem from demographic factors such as gender, age, race/ethnicity, educational attainment, and income. Some of the demographic differences are stark; for example, there is almost a 60-year age difference between the youngest (18-29 years) and oldest (age 65+) users of Snapchat and Instagram, while the same difference for Facebook is 20 years. Rural users of Facebook lag urban users by only 3% (67% versus 70%, in 2021), but the same gap is larger for LinkedIn (15% rural and 30% urban). Race/ethnic differences exist as well with African American and Hispanic use of Facebook slightly exceeding Whites but Whites outpacing Hispanics and African Americans in LinkedIn usage. All these differences bear resemblance to demographic, social, and economic differences that mark the digital divide in the United States.

The COVID-19 pandemic has shone fresh light on the digital divide in the United States. The digital divide has been defined as “the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies and to their use of the Internet for a wide variety of activities” (OECD, 2011). As socioeconomic differences between different groups are known to influence the digital divide in the United States, it is critical to ask – are similar factors playing a role in shaping the digital lives and habits of social media users? This is particularly relevant as the spectrum of internet use worldwide is diversifying. In the US, people are using the internet for online communication, electronic commerce, streaming entertainment, gaining access to the news and online education, remote working, and availing services such as tele-health and connecting remotely to household devices among an ever-broadening spectrum of online activities. This has resulted in shifting the attention of digital divide researchers and policy makers from studying the antecedents of internet access to purposeful internet usage (Author 2017; van Duersen and van Dijk, 2013). Similar to this expansive spectrum of online activities, social networking and overall social media use has purposes ranging from connecting with friends and family, meeting others with similar interests, connecting with support groups, social organizing, creating and sharing content to influence lifestyles and consumer preferences, accessing special perks provided by businesses, to signing up for events and supporting brands and organizations that align with one’s beliefs and values. Businesses and organizations interact with consumers via social media channels creating an ecosystem of self-expression, organization, sharing, communication, collaboration, and feedback. Use of social media is simply woven into the digital life of contemporary internet users. Therefore, similar to internet access and its purposeful use, it becomes essential to examine the antecedents of social media access and its purposeful use for a variety of activities.

The research objective of this study is to analyze spatial patterns, geographic disparities, and socioeconomic antecedents of social media access and purposeful social media (PSM) usage in U.S. counties. Specifically, we examine five popular social media platforms – social networks such as Facebook, Twitter, and LinkedIn, and media sharing platforms – Instagram and YouTube. In addition to the access of these 5 social media platforms, we also examine 12 indicators of purposeful use – from keeping in touch with friends and family, hedonic use such as playing online games, professional use such as networking with peers and contacts, sharing content such as photos and videos, to online influencing via supporting favorite brands and reviewing products and services. This focus on the spatial and socioeconomic analysis of purposeful social media use is a distinguishing feature of this work. It follows up on a prior study

(Pick, Sarkar, and Rosales, 2019) in which geographic patterns and socioeconomic determinants of three social networks (Facebook, Twitter, and LinkedIn) were examined. This study is different – in addition to two more social media platforms, a dozen indicators of PSM use are also analyzed.

The research questions are –

1. What are the spatial patterns of social media access and PSM use among US counties, as analyzed by geographic mapping and cluster analysis?
2. To what extent are spatial patterns of social media access and PSM use in U.S. counties influenced by spatial agglomeration?
3. What are the associations of demographic, socioeconomic, and social capital factors with social media access and PSM use in the U.S. counties?
4. What are the policy implications of such associations?

The study is conducted at the US county level. The county is an important administrative unit that lies below states in the hierarchy of US census geographic entities. According to the National Association of Counties, US counties are one of America's oldest forms of government dating back to the 1600s. While the structuring of county governments is diverse and basic roles and responsibilities of county governments are established by the states, counties administer their own economy, education, justice and public safety, healthcare, social services, transportation, and other services that directly impact the county residents. During the COVID-19 pandemic, the role of counties has especially come to the forefront in matters related to public health. In addition, counties are often responsible for implementing a broad array of federal, state and local programs including those in telecommunications. The choice of US county as the geographic unit of analysis is another distinguishing feature of this study. It provides a rich, nuanced view of disparities in social media access and purposeful use, especially when contextualized relative to a county's location and its surroundings.

The remainder of this paper is organized into sections on Literature Review of broader county-level studies focused on the digital divide in the U.S., conceptual model of social media access and purposeful use, spatial patterns of social media access and PSM use, regression findings, policy implications, limitations, and conclusions.

LITERATURE REVIEW

As social media has risen in importance in the U.S. and elsewhere, it became apparent that its use, purposes, and societal impacts vary by geographic location. What people use it for and how the locational environment is associated with particular uses have become important questions. So far, studies of the geographic aspects of social media use have been limited but are growing, especially as more data become available to researchers. This section will first review several prior studies on the spatial patterns of and socioeconomic associations with social media in the US. It next will talk about the importance of understanding social, economic, and political correlates of social media, with detailed review in the next section The section concludes by indicating the development in the literature of the concept and importance of purposeful use of technologies, as compared to access and simple use.

The spatial patterns and associations of social media at the state level in the US examined influences of 14 socio-economic, demographic, innovation, social capital, and societal openness variables on the eight measures of ICT access and utilization including two for social networking, i.e. Facebook use and Twitter use (Pick, Sarkar, and Johnson, 2015). Finding for social media indicated positive effects of social capital, with inverse effect of Hispanic percent for Facebook and inverse influence of college education for Twitter, while in less urban states Asian ethnicity related to Facebook use.

Research comparing profiles of social media uses and socioeconomic correlates for the US and Britain compared findings for surveys in Britain (OxIS) and US (Pew Research Center) that characterized the two sets of Twitter and non-Twitter users finding many differences, for instance Twitter users had greater exposure to college in the US, while British Twitter users had a much higher proportion of students. This cautions on making comparisons between nations or between groups in different settings. Unsurprisingly, the determinants of Twitter use differed in multivariate analysis, with the major British factors being high income, education, and retirement, compared to very high income, unemployment, and non-white race for the US. The study cautioned about applying them based on unrepresentative groups and mentioned that their use in business seems more appropriate.

A study some similar to the present one focused on Facebook, Twitter, and LinkedIn use, but not purposeful use, in 3,109 US counties in the lower 48 states in the years 2010-2012 (Author, 2019). The study identified demographic factors, service occupations, ethnicities, and urban location were the important determinant and that the three social media types had different spatial distributions nationally, with for example Twitter use heavy in southern California and Mississippi, while Facebook users had the highest intensities in the Rocky Mountain states. The study differs from the present one by being 8 years earlier and only analyzing general use or non-use of social media, rather than focused use such as for healthcare.

Studies by the Pew Research Center provide insight into socioeconomic differences in social media trends in the US, emphasizing the remarkable growth in this media use over time (Pew Research Center, 2018, 2021a, 2021b). For instance, in 2021, the social media types in greatest use in the US are YouTube (81% of adults), and Facebook (69% of adults), with lesser use of Instagram, Pinterest, LinkedIn, Snapchat, and Twitter, yet all the forms show consistent growth over time, with the exception of Facebook, levelling off in 2016 and Snapchat and Twitter, declining from 2018-2021. The social media nationally show differences by age, education, and race/ethnicity, and frequency of use, although with a lot of variation between types. For instance, Hispanic and Black Americans have 15% stronger use of Instagram than whites, while LinkedIn is used by half of those with college degrees, but only by one tenth of high school educated people (Pew Research Center 2021). The Pew studies inform the present study about up-to-date intensities of national use, but exclude geography and purposeful social media uses.

The present study is distinguished from prior ones by focusing on purposeful use of social media. This concept was early on referred to as a resources and appropriation theory (van Dijk, 2013, 2014). It may be thought of as a succession of stronger appropriation of social media, starting with (1) motivation to use it, followed by (2) physical infrastructure and material, (3) development of digital skills, (4) simple use, and (5) purposeful use for particular applications. The present study moves far along in this succession, past simple use/non-use, and centering on purposeful use. This connotes users who have a greater sense of target activities for their social media use. In educational research, this has been conceptualized as moving from access to educational technologies to specific academic outcomes such as outcomes involving creativity and innovation, social skills, cross-cultural skills, or leadership skills (Warschauer and Matuchniak, 2010). More recently, the educational skills necessary for purposeful outcomes have been categorized as operational, formal, information, communication, content-creation, and strategic (van Dijk, 2020).

For the present research, 12 purposeful uses of social media are the dependent variables of the analysis. They are studied relative to their determinants and to their spatial tendencies. The next section will build from the literature review by introducing the study's conceptual model and justifying its components.

CONCEPTUAL MODEL OF PURPOSEFUL SOCIAL NETWORKING

The conceptual model used, shown in Figure 1, is the Spatially Aware Technology Utilization Model (SATUM). SATUM proposes associations of socioeconomic, demographic, economic, market structure, infrastructure, social capital, and innovation factors with factors of Information and Communication Technology (ICT) access, use, and purposeful use. At the same time, it includes spatial analysis of the spatial distribution of the factors including the descriptive patterns, and clusters of high levels or low levels of utilization factors (sometimes referred to as "hot spots" and "cold spots").

It also includes measurement of spatial autocorrelation, i.e. the overall types of spatial agglomerations of factors, ranging from complete separation geographically of all the high values and all the low values to random distribution of values, to the other extreme of high values always being surrounded by low values and vice versa. Spatial autocorrelation usually tends towards the former type of agglomeration, due to Tobler's first law of geography, which indicates that like valued point location for a variable tend to agglomerate with other like-valued points (for more explanation of spatial autocorrelation, see Longley et al., 2015).

The SATUM model as applied in the present study combines exploratory analysis of five types of social network access along with analysis of purposeful social network use variables. The study's model (see Figure 1) indicates 10 independent variables in demographic, economic, and social bonding categories that influence five types of social media variables – on the upper right, and 12 types of purposeful use of social media in four groups – on the lower right. The specific geographic, multivariate, and clustering methods are detailed in the next section. This section provides justification of the choice of independent variables and categories of dependent variables, based on prior literature and on author reasoning.

Starting with the independent variables and the first group of demographic factors, *working age population* has been found be important in studies of the US at the state level (Pick, Sarkar, and Johnson 2015) and in Japan at the prefecture level (Nishida, Pick, and Sarkar, 2014). We reason that the working age population in the US has increasing business needs to utilize social media, adding to greater personal use as well. We posit that working age population is associated with social media use among types and with purposeful use.

Young dependency ratio indicates the ratio of children to the working age population, so in some sense represents age effects. In a prior study of social media use in the US, young dependency ratio was a consistent determinant of Facebook, Twitter, and LinkedIn use (Author, 2019). On the other hand, for Facebook use in the U.S. in April of 2021, social media use was substantially lower for teens than for working age population, which connotes a possible shift to an inverse relationship with social media (Statista, 2021). The posited change from the prior cited study in 2010-2012 is that today millennials are moving significantly into the workforce and as well social media is used more by working age people, so the social media impact of young dependency ratio shifts to a posited inverse effect. Hence, we propose that young dependency ratio has an inverse effect with use of social media types and purposeful use.

Ethnic and racial diversity has been of importance in geographical studies of technology utilization in the US (Grubesic, 2006; NTIA, 2011; Pick, Sarkar, and Johnson 2015), including analysis of its associations with social media (Pick, Sarkar, and Rosales, 2019; Campos-Castillo and Laestadius, 2020). For broadband in the US, both competition and availability were found to have association in the US for census tracts, but Hispanic households were found to have reduced broadband use (Grubesic, 2006). Similarly, for US Census data in 2021, Asian American percent was associated with higher broadband use, while percent Hispanic was related to lower broadband us (NTIA, 2011). For social media in 2010-2012, Asians had higher use for Twitter and LinkedIn, while Hispanics had lower Facebook use; Black had lower Facebook use but higher Twitter use. Research on differences among racial/ethnic categories for posting Covid-19 content on social media (Campus-Castillo and Laestadius, 2020) indicated that Black respondents were more likely than White respondents to report they posted covid-19 content on social media. Among men, the stronger postings went beyond Black respondents, and also applied to Latinos and other ethnic groups.

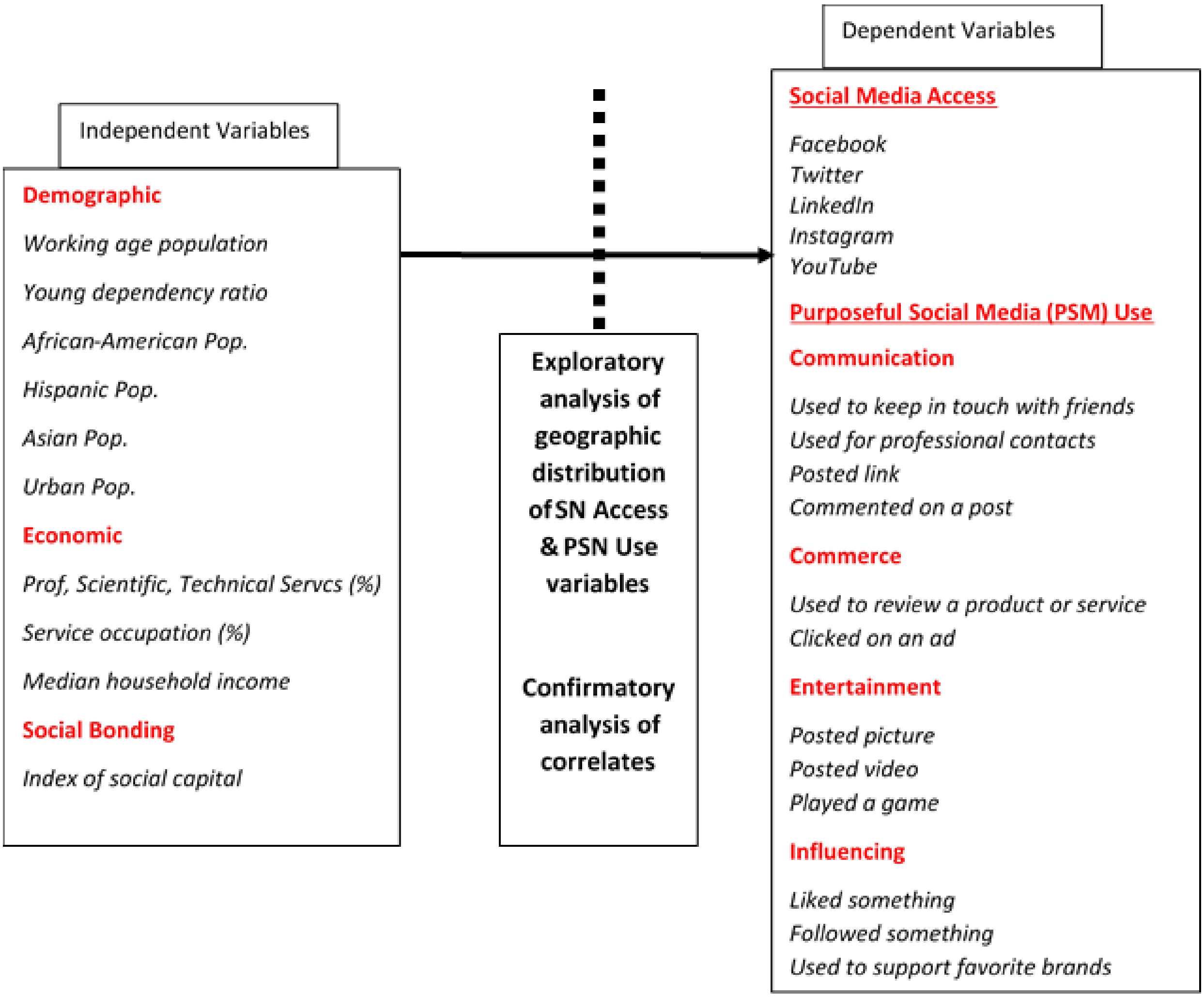
Altogether, the findings on race and ethnic associations indicate varied findings from study to study. Perhaps the only generalization is that Asian respondents tend to be associated with a wide varieties of technologies including social networking. In the present study, we posit that Asians will generally show positive effects across the categories of purposeful use, while Hispanics will mostly show inverse effects, and African Americans will be having reduced social media use, with the exception of higher use for Twitter and for health-related purposeful use.

Proportion of urban population has been found to be consistently positive in many studies of technological differences for geographies. Urban location found to be associated with higher broadband by zip codes across the US (Grubesic, 2005), while metropolitan areas in the US had higher broadband and mobile use than non-metropolitan areas (Prieger, 2013). For Facebook, Twitter, and LinkedIn in the US, urban location was consistently positive. We propose that urban location is associated with social media use across types of social media and is also positive for purposeful use.

The proportion of the *workforce in professional, scientific, and technical services* was positive in a study of influences on ICT variables at the county level (Azari and Pick, 2005), although it had no effects for social media variables in 2010-2012. We reason that the more varied and expansive uses of social media in 2018 would point to greater need for this workforce both as developers and users. We posit it is associated with social media uses among types of media and also for purposeful use of social media.

We reason *service occupations* would be related to social media uses because those occupations have more need to communicate and market by social media. In a study of hundreds of European Union regions in 27 countries, service employment was significant in its relationship to an ICT factor analysis score. The study pointed to services as essential due to their historical early role in the internet and the benefits were perpetuated (Vincente and Lopez, 2011). Several other studies of the US have shown positive effects. In a study of social media in 2010-12, services employees had no association with social media for the nation, but was inverse in effect for Twitter and LinkedIn in micropolitan areas and likewise for LinkedIn in rural areas, with the latter effects unexplained. We posit that service occupations has a positive effect on use by social media types and on purposeful uses.

Figure 1: Conceptual model of social media access and purposeful usage



Median household income: There are numerous studies in the literature showing positive effects of income on a variety of technology factors. Since it is one of the most supported variables for technologies, we posit it is related to social media use by types of social media and related to purposeful use of social media.

Dependent variables on purposeful use: The choice of dependent variables was drawn from a larger group of purposeful uses, and was rationalized based on the new interest in purposeful uses described generally earlier in this section (van Dijk, 2014, 2020; Warschauer and Matuchniak, 2010). We balanced the purposeful uses in four major categories we identified for purposeful use, namely Communication, Commerce, Entertainment, and Influencing.

In summary on the SATUM model and the justifications of specific choices of variables for this research, we found the SATUM model the most relevant to incorporating spatial distributions and agglomerations, as well as consideration of multivariate influences on social media dependent variables. We based the choice of variables on a combination of literature on the ICT digital divide and on a limited number of studies of the geographies in the US of social media use, as well as author reasoning related to the present study.

METHODOLODY

For the study, descriptive statistics for all 17 dependent variables and the 10 independent correlates were first computed. Correlation analysis was conducted next for pairwise correlations between independent variables to screen for multi-collinearity (Field, 2009, Hair et al., 2010). An example of an independent variable screened for multi-collinearity is educational attainment (measured by college education per capita) because of its strong positive correlation with median household income (Pearson Correlation Coefficient of 0.737, significant at the .05 level). Next, the dependent variables are mapped using a Geographic Information System (GIS). GIS mapping shows the spatial distribution of the study's dependent variables (Social Network access and PSM use) in US counties and provides important visual cues and insights about agglomeration of social networking activity in US counties. GIS mapping also helps to observe outlying values for the variables and points to counties with unduly high or low values of social network access and PSM use. In fact, spatial mapping helped identify 33 such counties with unduly high values of the dependent variable. Upon further examination, the normalized values of the dependent variables were found to be high for such counties all of which were sparsely populated often in rural areas. These 33 counties were excluded from the dataset leaving a robust set of 3,076 counties in the lower-48 states of the United States. In addition, when independent variables of interest such as age structure, urbanization, race/ethnicity, occupation are overlayed on the layers representing dependent variables in a GIS, spatial interactions between the dependent and independent variables can be explored.

Next, K-means cluster analysis, an essential data mining method is applied to the county dataset to assign counties to clusters that are similar in terms of their access of social networks (5 dependent variables) and also similar in terms of their PSM usage (12 dependent variables). In this study, after trial-and-error, counties are assigned to k = 5 clusters ranging from lowest to highest for social network access and PSM use. Two sets of k-means (k=5) clusters are obtained – one in which counties are assigned from lowest to highest social network access, and another in which counties are assigned from lowest to highest PSM use. Cluster centers are computed, and each cluster is characterized in terms of demographic, economic, and social attributes (the study's 10 independent variables). This yield important insights about the demographic and socioeconomic differences between the clusters. The clusters are also mapped in a GIS showing how clusters of low, moderate, and high social networking access and PSM use are spatially distributed in counties spanning the lower-48 states.

A key question is this study is whether social network access and PSM use in US counties shows statistically significant patterns of agglomeration of high and low values, or are such patterns spatially randomly distributed. We diagnose spatial autocorrelation – a common problem plaguing many technology diffusion datasets using Moran's I test statistic. The Moran's I test is inferential; the null hypothesis is that the values of a variable are randomly distributed spatially. The Moran's I test statistic ranges in value between –1 and +1. For example, Moran's I statistic value close to 0 for a dependent variable (social network access or PSM use) would indicate random spatial distribution of the variable among US counties. On the contrary, values close to –1 and +1 indicates 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 diagnosis is done by 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). In addition to spatial correlation analysis, local indicators of spatial association (LISA) based cluster-and-outlier analysis (Anselin, 1995) is also applied to identify statistically significant hotspots, coldspots, and outliers of social network access and use. The insights obtained from LISA-based classification of counties and LISA mapping shows hotspots and coldspots to be often associated with the extent of urbanization in counties. The outlier counties are of two types; high-low outliers are counties that have a high level of social network access or PSM use whose neighboring counties have low levels of social networking activity, whereas low-high outliers are just the reverse. Mapping of hotspots, coldspots, and outlier counties provoke further examination of such counties to uncover underlying reasons of such spatial patterns.

Lastly, posited associations between the 17 dependent variables and 10 independent correlates are examined using Ordinary Least Squares (OLS)-based regression analysis. OLS regressions were conducted in stepwise fashion, allowing in only those of the independent variables with significance levels of equal or less than 0.05. As an additional test of multi-collinearity, the variance inflation factor (VIF) was computed for each independent variable. We utilized the common cut-off of 5 or greater for VIF to be of concern (Myers, 1990) and no multicollinearity problems were detected. Three diagnostic tests were administered to ensure that regression assumptions were met. Joint Wald Statistic is a test of the joint significance of several coefficients of individual independent variables (Wald, 1943). The Koenker (BP) Statistic Test is a test for heteroscedasticity, i.e. the variance of the residuals is not constant (Lyon and Tsai, 1996). The Jarque-Bera Statistic is a goodness-of-fit test of whether sample data, in this case regression residuals, have skewness and kurtosis that correspond to a normal distribution (Jarque and Bera, 1980). Additionally, regression residuals were tested for the presence of spatial bias using Moran's I test statistic. Model relationships which result in spatially random errors are regarded as valid. If errors in the model fit are spatially autocorrelated, it implies that the geographic forces are exogenous to the conceptual model. In case Moran's I testing indicates that regression residuals are not spatially randomly distributed, regressions results have to be treated with caution.

In summary, the methodology of the study is comprised of GIS mapping of dependent variables, screening for spatial bias and spatial patterns of social network access and use using geostatistical methods, and OLS regressions. The application of geostatistical methods such as spatial autocorrelation (Moran's I) and LISA analysis – while common for health, social, and economic outcomes such as incidences of diabetes, poverty, crime, etc. is relatively uncommon for the study of technology adoption and utilization and is a novel feature of this study.

DATA

Data for various independent and dependent variables used in the study were collected from multiple sources. Data for 7 out of the 10 independent variables (age structure, race/ethnicity, urban population, and median household income) were sourced from the U.S. Census Bureau's American Community Survey. 5-year estimates (2015-2019) centered on 2017 was used for these 7 independent variables. It includes data collected between January 1, 2015 and December 31, 2019. According to the U.S. Census Bureau, 5-year estimates are the most reliable, based on the largest sample size, and include data all areas, unlike 3- and 5-year estimates that have population cutoffs of 20,000 or more and 65,000 or more respectively. Data for two of the occupation-related independent variables – professional, scientific, and technical services employment, and service sector employment were collected from Esri's Business Analyst software for the year 2020. To estimate such data, Esri combines its proprietary

Tapestry Segmentation data with data obtained from various surveys such as the MRI Survey of the American Consumer, and Doublebase from MRI-Simmons which is an integration of information from four consumer surveys. More details about Esri's methodology to collect such data can be found in Esri (2020). The index for social capital is an amended version of a social capital index designed as part of the "The Geography of Social Capital in America" project of the United States Congress Joint Economic Committee. The county-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. The county-levels social capital index data is for the year 2018, and more details about the index and its research design methodology can be found in U.S Senate (2018).

Data for all 17 dependent indicators of social media (Facebook, Instagram, LinkedIn, YouTube, and Twitter) access and PSM use for communication, commerce, entertainment, and influencing were also obtained from Esri's Business Analyst software for the year 2020. As mentioned earlier, such data are estimated from a variety of surveys on consumer preferences and consumer use of technology. It is important to note that the surveys based on which social network access and use data were estimated ran through Spring 2019 (Esri, 2020). Therefore, the data does not reflect changes in consumer behavior attributable to coronavirus (COVID-19) pandemic.

Since data for 8 out of the 10 independent variables are from 2017 and 2018 and data for all 17 dependent variables are from 2020, time simultaneity has been maintained. While data for the two occupational independent variables (PSTS and service sector employment) are from 2020, such employment data changes has not been prone to significant changes, at least prior to the COVID-19 pandemic. So having data for these two independent variables for the year 2020 does not affect the study design.

Finally, data for all variables is for counties in the lower-48 states of the United States. Counties and administrative units in Alaska and Hawaii are excluded from the study due to incompleteness. As mentioned earlier, a handful of sparsely populated counties randomly distributed in the lower-48 states were excluded since the dependent variable values were identified to be outliers. All variables were converted to per capita whenever possible; variable definitions, sources, and descriptive statistics (n = 3,076 counties) of the dependent and independent variables are in Table 1.

Table 1: Definitions and Descriptive Statistics of Dependent and Independent Variables

	Dependent Variable	Definition	Year	Source	MAX*	MIN*	Mean*	SD*
Social Network Access	Facebook	Facebook access, last 30 days (% of Total Pop.)	2020	Esri Business Analyst (BA)	0.33	0.70	0.50	0.05
	Instagram	Instagram access, last 30 days (% of Total Pop.)	2020	Esri BA	0.08	0.51	0.20	0.05
	LinkedIn	LinkedIn access, last 30 days (% of Total Pop.)	2020	Esri BA	0.02	0.27	0.06	0.04

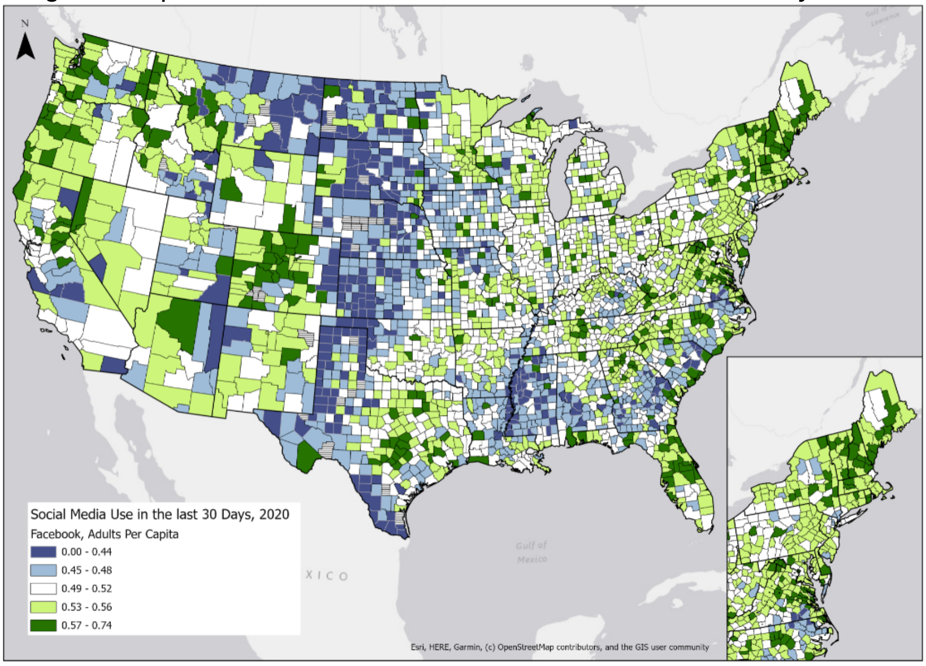
	YouTube	YouTube access, last 30 days (% of Total Pop.)	2020	Esri BA	0.01	0.19	0.03	0.02
	Twitter	Twitter access, last 30 days (% of Total Pop.)	2020	Esri BA	0.03	0.27	0.09	0.03
Purposeful Social Network (PSM) Use	Posted picture	Posted picture, last 30 days (% of Total Pop.)	2020	Esri BA	0.20	0.55	0.34	0.04
	Posted video	Posted video, last 30 days (% of Total Pop.)	2020	Esri BA	0.08	0.25	0.15	0.02
	Posted link	Posted link, last 30 days (% of Total Pop.)	2020	Esri BA	0.03	0.23	0.09	0.02
	Commented on a post	Commented on a post, last 30 days (% of Total Pop.)	2020	Esri BA	0.21	0.53	0.36	0.04
	Played a game	Played a game, last 30 days (% of Total Pop.)	2020	Esri BA	0.09	0.25	0.16	0.02
	Liked something	Liked something, last 30 days (% of Total Pop.)	2020	Esri BA	0.19	0.48	0.32	0.03
	Followed something	Followed something, last 30 days (% of Total Pop.)	2020	Esri BA	0.09	0.39	0.17	0.03
	Clicked on an ad	Clicked on an ad, last 30 days (% of Total Pop.)	2020	Esri BA	0.05	0.25	0.13	0.02
	Used to keep in touch with friends	Used to keep in touch with friends, last 30 days (% of Total Pop.)	2020	Esri BA	0.22	0.46	0.34	0.03
	Used to review a product or service	Used to review a product or service, last 30 days (% of Total Pop.)	2020	Esri BA	0.01	0.10	0.04	0.01
	Used for professional contacts	Used for professional contacts, last	2020	Esri BA	0.03	0.13	0.06	0.01

		30 days (% of Total Pop.)						
	Used to support favorite brands	Used to support favorite brands, last 30 days (% of Total Pop.)	2020	Esri BA	0.02	0.08	0.05	0.01
	Independent Var.							
Age Structure	Working Age Pop.	Pop 15 - 64 / Total Pop	2017	US Census	0.40	0.82	0.63	0.04
	Young Dependency Ratio	Pop. 0-15 / Total Pop. 16-64	2017	US Census	0.06	0.59	0.29	0.05
Race / Ethnicity	African American	Af-Am Pop / Total Pop	2017	US Census	0.00	0.87	0.09	0.15
	Asian	Asian Pop / Total Pop	2017	US Census	0.00	0.35	0.01	0.02
	Hispanic	Urban Pop / Total Pop	2017	US Census	0.00	0.99	0.09	0.14
Geography	Urban	Hisp. Pop / Total Pop	2017	US Census	0.00	1.00	0.41	0.32
Economic: Income & Employment	Median Household Income	Median Household Income	2017	US Census	19,264.00	129,588.00	49,530.27	12,994.87
	Professional, Scientific & Technical Services (PSTS) Employment	PSTS Employment Age 16+ / Tot Pop 16 + Yrs	2020	Esri BA	0.00	0.32	0.02	0.02
	Service Sector Employment	Service Sector Employment Age 16+ / Tot Pop 16 + Yrs	2020	Esri BA	0.03	0.51	0.17	0.05
Bonding	Social Capital	Modified Social Capital Index	2018	Joint Economic Committee, US Congress	0.00	7.18	4.31	0.98
		** n = 3,076 counties*						

FINDINGS: SPATIAL PATTERNS OF SOCIAL MEDIA ACCESS AND PURPOSEFUL USE

GIS mapping of all 17 dependent indicators yielded descriptive visualizations of social media access and purposeful use in US counties. Such visualizations shows spatial patterns of social media access and activity nationwide and provide insights on possible agglomeration of high versus low areas of access and use which have to be statistically tested for agglomeration effects. Figure 2 shows that Facebook use among American social media users was high (53 – 74% of adult users, last 30 days) in bicoastal, largely urban, metropolitan areas in California, Florida, large parts of the North-eastern seaboard, the Great Lakes region, as well as the Rocky Mountain states. Low levels of social networking via Facebook was more common in the rural south as well as the prairie states. Similar spatial distributions for Facebook were observed for a prior study (Pick, Sarkar, and Rosales, 2019). While similar patterns were also observed for other social networking platforms such as LinkedIn and Twitter, the intensity of use (measures by adults per capita) was significantly lower than Facebook reiterating that Facebook continues to remain the dominant social networking platform. The same was true for media platforms such as YouTube and Instagram. Correspondingly for the 12 PSM use dependent variables, spatial distributions were largely similar to the social network access variables. One difference was counties with highest levels of social media access often have high to moderate levels of social media purposeful usage (see figure 3 for example, which shows the spatial distribution of social media use for content sharing – posting a video). It is important to note that the maps in Figures 2 and 3 have different category breaks making comparable interpretation of the various categories (low, moderate, high) hard to make.

Figure 2: Spatial distribution of Facebook access in last 30 days, 2020



K-means cluster analysis (K = 5) shows lowest levels of social media access in 769 and low levels in another 1,383 counties, accounting for over 2 out of 3 counties (approximately 70%) in the lower-48 states. High-highest levels of social media access is found in 323 counties, which is about 10% of all counties included in the study (Table 2). Among social networks, counties in the highest cluster were found to have 5.57 times LinkedIn access compared to those in the lowest cluster. Corresponding ratios for YouTube and Instagram were 4.90 and 2.73 indicating the extent of disparity between highest and lowest access clusters. Counties in cluster 5 (highest access) have higher working age population and lower young dependency ratios compared to those in cluster 1 (lowest access), indicating the presence of adults in the household who are more likely to access social media platforms for networking, content sharing, and other purposes. These counties were located in counties with significantly higher urban population per capita, reported higher median household income, and had higher proportions of their populations in professional, scientific, and technical services occupations and service sector occupations (Table 3). Clearly, such counties are prosperous with higher levels of educational attainment that catalyze higher levels of social media access. Geographically, counties in the highest and high access clusters are in the Rocky Mountain states, particularly Colorado, Boston-Washington, D.C. megalopolitan area in the Northeast, coastal California as well as the Pacific Northwest. Lowest access counties are in the prairie states, rural south, and Appalachia region while low access counties are in the Midwest including the Great Plains region among others (Figure 4).

Figure 3: Spatial distribution of PSM use (Posting a Video), last 30 days, 2020

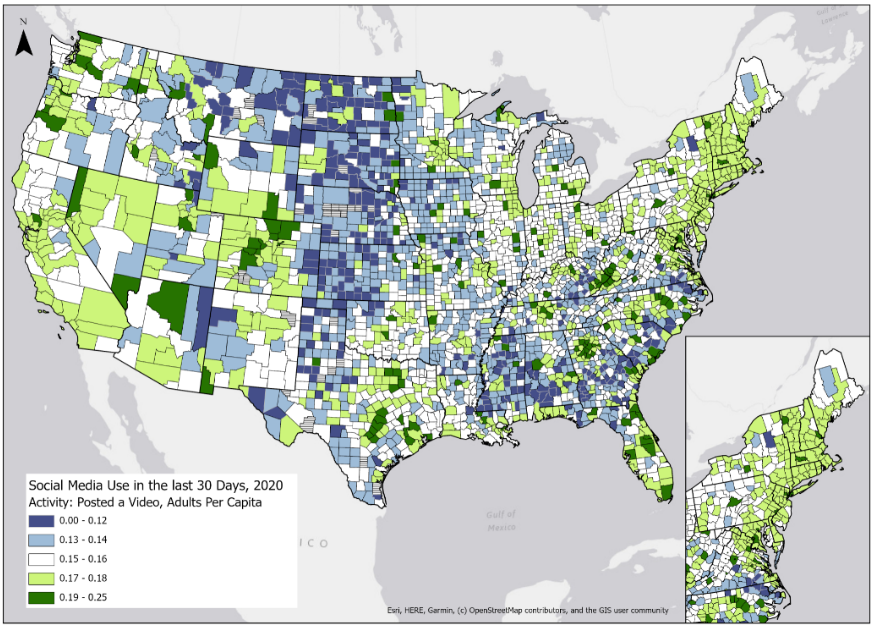


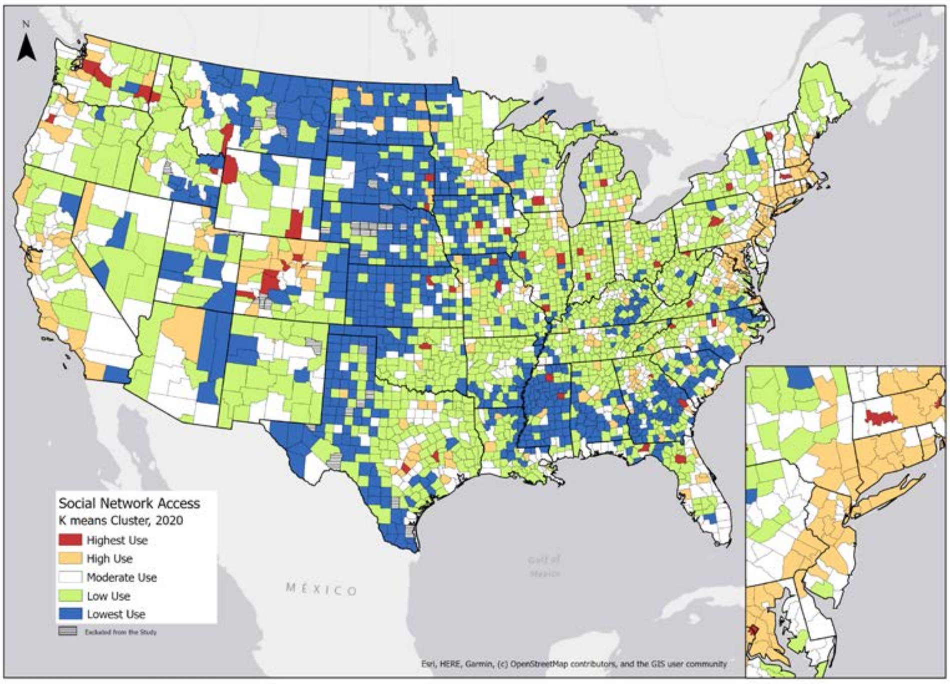
Table 2: Cluster Centers – Social Media Access

	Lowest Access	Low	Moderate	High	Highest Access	MAX	MIN	Ratio = Highest / Lowest
Facebook	0.44	0.51	0.54	0.55	0.61	0.61	0.44	1.37
Instagram	0.14	0.19	0.24	0.29	0.38	0.38	0.14	2.73
LinkedIn	0.03	0.05	0.09	0.13	0.19	0.19	0.03	5.57
YouTube	0.02	0.02	0.04	0.07	0.10	0.10	0.02	4.90

Twitter	0.07	0.08	0.11	0.14	0.20	0.20	0.07	2.94
n	*769*	*1383*	*601*	*262*	*61*			

Table 3 Cluster Characteristics – Social Media Access						
		K=1	K=2	K=3	K=4	K=5
	Independent Var.	Lowest SN Access	Low	Moderate	High	Highest SN Access
Age Structure	Working Age Pop.	0.62	0.63	0.65	0.68	0.73
	Young Dependency Ratio	0.32	0.29	0.29	0.28	0.20
Race / Ethnicity	African American (%)	0.14	0.07	0.09	0.11	0.09
	Asian (%)	0.00	0.01	0.02	0.05	0.05
	Hispanic (%)	0.10	0.07	0.11	0.11	0.09
Geography	Urban (%)	0.27	0.34	0.64	0.80	0.75
Economic: Income & Employment	Median Household Income ($)	44,122.83	45,516.09	55,840.53	70,419.08	56,818.80
	PSTS Employment (%)	0.01	0.01	0.03	0.05	0.05
	Service Sector Employment (%)	0.16	0.15	0.20	0.26	0.27
Bonding	Social Capital	4.44	4.24	4.18	4.52	4.45

Figure 4: K-means clusters, Social Network Access (Facebook, Instagram, LinkedIn, YouTube, Twitter), 2020



For PSM use, a lower proportion of counties (48.15%) are in the lowest and low use clusters, while a higher proportion (16.78%) are in the high and highest use clusters. This is a marked difference compared to the cluster distribution of counties for the 5 social media access variables. Also, the intensity of PSM use in the highest versus lowest use clusters in comparatively more balanced with ratios ranging between 1.36 and 2.89 (Table 4). This shows that among those who access social media platforms, usage in more balanced for purposes such as social networking, sharing and posting content, commerce, and influencing purposes. The counties in the highest PSM use clusters have similar socioeconomic differences (Table 5) compared to those in the low and lowest PSM use clusters and spatial patterns (shown in Figure 5) are also similar.

Table 4: Cluster Centers – Purposeful Social Media Use

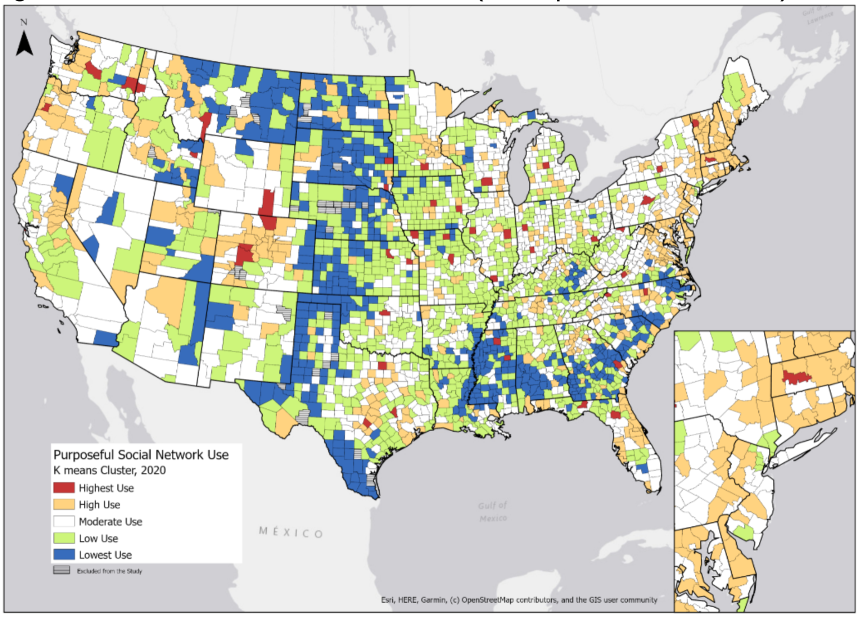
	Lowest PSM Use	Low	Moderate	High	Highest PSM Use	MAX	MIN	Ratio = Highest / Lowest
Posted picture	0.27	0.32	0.35	0.39	0.47	0.47	0.27	1.71
Posted video	0.12	0.14	0.16	0.17	0.21	0.21	0.12	1.79
Posted link	0.06	0.08	0.10	0.12	0.17	0.17	0.06	2.89
Commented on a post	0.29	0.34	0.37	0.41	0.46	0.46	0.29	1.59
Played a game	0.14	0.16	0.17	0.17	0.20	0.20	0.14	1.46
Liked something	0.26	0.30	0.33	0.36	0.41	0.41	0.26	1.55
Followed something	0.13	0.16	0.18	0.21	0.29	0.29	0.13	2.27
Clicked on an ad	0.10	0.12	0.13	0.16	0.20	0.20	0.10	2.09
Used to keep in touch with friends	0.30	0.33	0.35	0.37	0.40	0.40	0.30	1.36
Used to review a product or service	0.04	0.04	0.04	0.05	0.07	0.07	0.04	1.80
Used for professional contacts	0.05	0.05	0.06	0.07	0.10	0.10	0.05	1.89
Used to support favorite brands	0.04	0.04	0.05	0.05	0.06	0.06	0.04	1.40
n	*506*	*975*	*1079*	*464*	*52*			

Table 5: Cluster Characteristics – PSM Use

		K=1	K=2	K=3	K=4	K=5
	Independent Var.	Lowest PSM Use	Low	Moderate	High	Highest PSM Use
Age Structure	Working Age Pop.	0.62	0.62	0.64	0.66	0.73

	Young Dependency Ratio	0.33	0.30	0.28	0.27	0.20
Race / Ethnicity	African American (%)	0.18	0.08	0.07	0.08	0.08
	Asian (%)	0.00	0.01	0.01	0.03	0.05
	Hispanic (%)	0.12	0.09	0.08	0.09	0.07
Geography	Urban (%)	0.30	0.32	0.47	0.65	0.73
Economic: Income & Employment	Median Household Income ($)	42,847.80	45,295.75	50,446.80	63,444.87	50,774.12
	PSTS Employment (%)	0.01	0.01	0.02	0.04	0.03
	Service Sector Employment (%)	0.16	0.16	0.17	0.23	0.25
Bonding	Social Capital	4.19	4.34	4.24	4.52	4.46

Figure 5: K-means clusters, PSM Use (12 dependent variables), 2020



Cluster-and-outlier analysis, based on Local Indicators of Spatial Association (LISA), a spatial statistical method (Anselin, 1995) reveals statistically significant hotspots, coldspots, and outliers of social media access and PSM use in the study area. While the hotspots and coldspots largely conform to urban-rural differences, outlier counties (county with high level of social media activity surrounded by counties with low levels of such activity, or vice-versa) have yielded interesting insights. For several dependent variables, such counties are found to comprise a university/college town or community within the county boundary or are located within the service are of an institution of higher education. An example is Buffalo County, Nebraska, whose county seat Kearney is home to a campus of the University of Nebraska. It is a statistically significant outlier with high levels of LinkedIn access and social network use for professional networking, but is surrounded by counties with low levels of such activity. This finding is being scrutinized more at the time of finalizing this manuscript.

REGRESSION FINDINGS: SOCIO-ECONOMIC DETERMINANTS OF SOCIAL NETWORK ACCESS AND PURPOSEFUL USE

Two sets of OLS regressions were conducted for the 5 social network access (Facebook, Instagram, LinkedIn, YouTube, and Twitter) and the 12 PSM use dependent variables.

As seen in Table 6, the dominant correlates of social network access are demographic variables such as urban population, age structure (young dependency ratio and working age population), and professional, scientific, and technical services (PSTS) employment, followed by service sector employment. Among race/ethnic variables, African American population has association with 3 of the 5 social network access variables (Facebook, Instagram, and Twitter), while the association of Hispanic population with 2 of the 5 variables is mixed. Both median household income and social capital are not found to have any association with this set of dependent variables. Overall, this set of OLS regressions explains 51% - 79% of the variation in the dependent variables.

Table 6: OLS Regression Findings, Social Media Access, 2020

Ind. Var. (Refer to Table 1)	Facebook	Instagram	LinkedIn	YouTube	Twitter
Working Age Pop.	0.174***	0.365***	0.232***		0.344***
Young Dependency Ratio	-0.354***	-0.095***	-0.14***	-0.226***	-0.102***
African American	-0.306***	-0.174***			-0.141***
Asian				0.235***	
Hispanic	-0.170***			0.125***	
Urban	0.264***	0.368***	0.310***	0.213***	0.257***
Median Household Income					
Professional, Scientific and Technical Services (PSTS) Employment	0.319***	0.315***	0.286***	0.259***	0.118***
Service Sector Employment	-0.169***		0.286***	0.329***	0.393***
Social Capital					
Variance Inflation Factor	2.392	1.658	2.312	2.538	2.329
Adjusted R^2	0.509***	0.708***	0.773***	0.788***	0.742***
Sample Size	3076	3076	3076	3076	3076
Regression Diagnostics					
Joint Wald Statistic	1711.168***	1324.166***	2874.981***	3117.313***	2996.311***
Koenker Statistic	181.701***	506.905***	511.752***	480.085***	382.931***
Jarque-Bera Statistic	1939.356***	12146.994***	11065.771***	31468.227***	8125.280***

Spatial Autocorrelation Analysis					
Moran's I of Dependent Var	0.391***	0.321***	0.339***	0.433***	0.258***
Moran’s I (Std. Resid)	0.218***	0.159***	0.208***	0.157***	0.152***
	* p < 0.05, ** p < 0.01, *** p < 0.001				

The positive associations of working age population, urban population per capita, and service sector occupation are all consistent with a prior study that examined socioeconomic determinants of social media access in US counties (Pick, Sarkar, and Rosales, 2019). Recent surveys (Pew, 2021) have shown that there are stark age-based differences in social network access among American adults with close to 80-84% of adults ages 18 to 49 using any social media sites, while only 45% of adults older than 65 years reported using social media. The positive association of urbanization with the adoption and use of the internet and other information and communication technologies (ICTs) has been well-documented in the digital divide literature (Pick, Sarkar, and Johnson, 2015; Chen, 2013). Further rural use of social media – particularly Instagram and LinkedIn trials urban use by 20% and 15% respectively (Pew, 2021). It must be noted though that there is relatively a negligible gap in internet use for social networking between Americans in urban (73.8%) versus rural (73.3%) areas in 2019 (NTIA, 2019). PSTS employment is positively associated with all 5 social network platform access and service employment with 3 platforms (LinkedIn, YouTube, and Twitter) indicating the propensity of those employed in legal services, accounting, tax preparation, bookkeeping, and payroll services, architectural, engineering, and related services, design services, computer systems design services, management, scientific, and technical consulting services, scientific research and development services, and advertising, public relations, and related services to access social networking platforms for communication, commerce, professional networking, and brand advertising purposes. Another explanation of the positive association of PSTS and service occupations with social network access may be the higher levels of educational attainment among those employed in these sectors as demonstrated by strong positive correlation between college education per capita with PSTS and services (.729 and .871 Pearson Correlation Coefficients, respectively, significant at .05 level). This finding further corresponds to the importance of professional and service occupations for payroll and receipts in most technology sectors for U.S. counties from 1997 to 2000 (Azari and Pick, 2005), and a prior study examining focused solely on Facebook, Twitter, and LinkedIn (Pick, Sarkar, and Rosales, 2019).

Compared to this prior study, the association of young dependency ratio (YDR) is reversed, and YDR is found to be negatively associated with social network access. In other words, holding the population 16-64 years constant, as the population 0-15 years increases, social network access for the 5 platforms – Facebook, Instagram, LinkedIn, YouTube, and Twitter decreases. This may be explained by the strong following of Instagram, Snapchat and TikTok among young adults 18-30 years old and the continuous migration of younger individuals (tweens and teens) to newer social media platforms such as Tiktok and Snapchat compared to Facebook (Auxier and Anderson, 2021). Among race/ethnicity, the inverse association of African American population per capita with Facebook, Twitter, and Instagram has some overlap with findings from a prior study (Pick, Sarkar, and Rosales, 2019). Among the independent correlates, median household income and social capital are found to have no association with any of the five social network access variables.

As shown at the bottom of Table 6, the Moran's I index of spatial autocorrelation decreases for each of the 5 dependent variables. For example, Facebook has a Moran's I value of 0.391 significant at the .001 level; the Moran's index for the regression residual of Facebook is 0.218, also significant at the .001 level, a decrease of 44.25%. The decrease in Moran's I ranges between 39% - 64% showing the conceptual model is able to account for spatial bias diagnosed in each of the 5 dependent variables.

Regression findings for the 12 PSM use dependent variables are in Table 7. The dominant correlates of purposeful social networking for communication, commerce, entertainment, and influencing are PSTS employment, urban population, race/ethnicity (African American and Hispanic population), and age structure (working age population and young dependency ratio). The directions of associations are found to be similar to the ones found for the previous set of regressions for the 5 social network access variables. For example, PSTS employment and urban population each have positive associations with the 12 PSM use dependent variables along with working age population, while young dependency ratio is negatively associated. An interesting difference is the rather dominant negative influence of African American and Hispanic populations per capita with 10 and 9 out of the 12 PSM use variables respectively. Another noticeable difference is the inverse association of median household income with 5 of the 12 PSM use variables. Among other differences, the role of service sector employment is rather muted compared to the first set of regressions, with this independent variable having mixed (both positive and negative associations) with only 4 of the 12 PSM use dependent variables. Similarly, Asian population per capita has mixed effects, while social capital is found to be negatively associated with 3 of the 12 variables. Overall, this set of regressions explain 39% - 67% of the variation in the PSM use dependent variables.

The rationale for positive associations of PSTS employment, urbanization, and working age population with PSM use stems from those residing in urban areas and employed in PSTS often have higher levels of educational attainment and tend to possess digital knowledge and skills to be prolific, sophisticated users of social media for a wide spectrum of digital activities spanning communication with friends and family, to professional networking and information dissemination, to engaging in electronic commerce and entertainment, to influencing by supporting brands, products, causes, and services. The rationale for the inverse association with young dependency ratio is likely to be similar to the explanation for the previous set of regressions (for the 5 social network access variables). The inverse associations of African American and Hispanic populations with PSM use is explained by the negative correlation of smartphone ownership with these race/ethnic groups (Pearson Correlation Coefficients of -.196 and -.115, significant at .05 level) but needs further research. Like the previous set of regressions, Moran's I of regression residuals of all 12 PSM use variables are lower than the Moran's index for the dependent variable itself, indicating that the proposed model accounts for spatial bias.

Overall, this set of findings of socioeconomic determinants of purposeful social network use are novel and have not been analyzed previously. While there are some similarities in both sets of regressions (for example, with respect to PSTS employment, urbanization, age structure), there are some differences as well (for example, with respect to race/ethnicity being more dominant for PSM use rather than social network access). The muted, inverse association of social capital with 3 indicators PSM use points to the fact that lack of digital access as well as lack of digital skills – traditional impediments to purposeful internet use, are not so much of an issue for social networking use.

Table 7: OLS Regression Findings, Purposeful Social Media Use, 2020

Ind. Var. (Refer to Table 1)	Posted picture in last 30 days	Posted video in last 30 days	Posted link in last 30 days	Commented on a post in last 30 days	Played a game in last 30 days	liked something in last 30 days	Followed something in last 30 days	Clicked on an ad in last 30 days	Used to keep in touch with friends	Used to review a product or service	Used for professional contacts	Used to support favorite brands
Working Age Pop.	0.306***	-0.108***	0.266***	0.117***		0.102***	0.398***		0.122***	0.332***	0.461***	0.369***
Young Dependancy Ratio	-0.233***	-0.108***	-0.164***	-0.332***	-0.325***	-0.374***	-0.121***	-0.325***	-0.406***			
African American	-0.234***	-0.251***	-0.307***	-0.364***	-0.150***	-0.432***	-0.320***	-0.297***	-0.354***	0.176***	0.274***	
Asian				-0.091***					-0.101***	0.138***	0.143***	0.102***
Hispanic	-0.08***		-0.127***	-0.167***	-0.258***	-0.096***	-0.149***	-0.101***	-0.123***		0.096***	
Urban	0.287***	0.357***	0.392***	0.267***	0.439***	0.194***	0.389***	0.275***	0.232***	0.197***	0.143***	
Median Household Income					-0.286***		-0.129***			-0.236***	-0.269***	-0.577***
Professional, Scientific and Technical Services (PSTS) Employment	0.271***	0.228***	0.210***	0.327***	0.194***	0.292***	0.234***	0.133***	0.218***	0.356***	0.267***	0.267***
Service Sector Employment			0.106***		-0.320***			0.379***		-0.312***		
Social Capital		-0.105***								-0.166***		-0.227***
Variance Inflation Factor	1.666	1.698	2.392	1.735	3.116	1.659	2.412	2.367	1.735	3.456	2.498	2.732
Adjusted R^2	0.583***	0.561***	0.664***	0.539***	0.393***	0.563***	0.601***	0.647***	0.470***	0.654***	0.660***	0.490***
Sample Size	3076	3076	3076	3076	3076	3076	3076	3076	3076	3076	3076	3076
Regression Diagnostics												
Joint Wald Statistic	1536.249***	1304.385***	1680.246***	1835.384***	781.549***	2013.121***	1450.707***	2541.083***	1572.614***	3917.666***	3569.953***	2648.803***
Koenker Statistic	238.823***	310.506***	537.123***	305.324***	463.957***	235.668***	547.128***	311.402***	137.338***	142.523***	196.992***	66.339***
Jarque-Bera Statistic	2373.921***	2640.992***	2506.839***	1554.860***	958.156***	1703.611***	2274.663***	1879.903***	1781.217***	3157.472***	2730.330***	1471.672***
Spatial Autocorrelation Analysis												
Moran's I of Dependent Var	0.362***	0.325***	0.331***	0.387***	0.425***	0.375***	0.281***	0.375***	0.339***	0.533***	0.348***	0.355***
Morans I (Std. Resid)	0.221***	0.178***	0.217***	0.233***	0.251***	0.184***	0.195***	0.189***	0.191***	0.176***	0.140***	0.131***
		* p < 0.05, ** p < 0.01, *** p < 0.001										

Before concluding this section, it is important to observe that Variance Inflation Factor (VIF) values for all 17 regressions are below the 5.0 threshold indicating that multi-collinearity is not a problem. Regression assumptions are largely met as evident from the values of the regression diagnostics. In addition, all coefficients in the regressions (results in Tables 6 and 7) are statistically significant at the .001 level. This is due to the large overall sample (n = 3,076 counties) of counties.

It is therefore essential to examine the magnitude of the standardized coefficients to determine which independent correlates have higher associations with social media usage. Further, given the large sample size fallacy, it is important to focus on effect size (Lantz, 2013; Sullivan and Feinn, 2012). According to Sullivan and Feinn (2012), the adjusted r-squared is an acceptable effect size index for studies that examine associations between variables. Adjusted r-squared values range between approximately 0.50 and 0.80 in this study with a couple of exceptions: playing a game and keeping in touch with friends – both PSM use variables. Such effect size is characterized as large (Sullivan and Feinn, 2012) indicating that large sample size fallacy (Lantz, 2013) has been addressed.

IMPLICATIONS OF FINDINGS

Both spatial analysis of geographical patterns as well as regression analysis show the association of urbanization with social media activity. This points to the well-known urban-rural divide for ICT adoption and diffusion. Such gaps point to the need for training programs targeted at rural users of social media. This can include introducing social networking platforms other than Facebook such as Twitter and LinkedIn to rural users and communicating the benefits of networking using such platforms. Beyond social networking platforms, other platforms that may appeal to rural users such as content sharing platforms such as YouTube and Instagram, as well as other types of social media platforms blogging (Tumblr and Wordpress), and collaborative websites (Wikipedia) can be introduced and their benefits explained.

County governments that have an influence on ICT policy can provide support for social media training and workshops for their citizenry, particularly those residing in rural communities and encourage the hiring of local ICT graduates (Kvasny and Keil, 2006), encourage and help with incentives for service sector industries in rural communities to make greater use of social media and perhaps even deploy social media analytics for business benefits. Regarding social media training, the governments need citizens to go beyond just completing courses or certificates, but to leverage the training for the “next step,” which might be further education or a closely-related job (Kvasny and Keil, 2006).

Other suggested actions may include –

1. Attracting more professional, scientific, and technical services and service sector employees to laggard communities and areas and encouraging existing service-oriented organizations in such areas to infuse social media in their operations;
2. Encouraging PSTS jobs for younger populations and families and stemming the outflow of the young population from impacted counties;
3. Encouraging attracting more professional, scientific, and technical workers, which may be challenging task given paucity of enterprises that would attract such workers;
4. County governments might also seek to broaden the purview of existing social capital by encouraging rural community organizations and human social networks to emphasize social

media knowledge and their purposeful use for building community knowledge, communal organizing of events and activities, and other forms of purposeful usage of social media.

CONCLUSION, LIMITATIONS, AND FUTURE WORK

This paper examines spatial patterns and socioeconomic influences on social media access and purposeful social media usage in US counties. Using the Spatially Aware Technology Utilization Model (SATUM), associations of demographic, economic, and social capital are posited with 5 indicators of social media access via Facebook, Instagram, LinkedIn, YouTube, and Twitter, and 12 indicators of PSM use. Dominant correlates of social media access are age structure, professional, scientific, and technical services occupation, and service sector occupation, and urbanization. In addition to these variables, purposeful use of social media for communication, commerce, entertainment, and influencing are found to be associated with race/ethnicity. Spatial patterns of social media access and use show agglomeration in terms of clusters. Based on K-means clustering, social media access levels are low in almost 7 out of 10 US counties, and high in 1 out of 10. In contrast, purposeful use of social media is low in half of the US counties in the study sample, and high in about one-sixth of US counties. Low levels of access and PSM use is found predominantly in the prairie states as well as in the South in counties with low levels of urbanization while higher levels are found in the Rocky Mountain states, Boston-Washington megalopolitan area, and coastal, urban areas. Outliers are often found in counties with large, public universities. Implications of these findings are discussed.

A theoretical limitation of the study stems from the lack of incorporation of attitude and motivational factors among social media users in the study's conceptual model. While SATUM is positioned to examine associations between social media access and use with socioeconomic antecedents while simultaneously factoring geographic differences and spatial bias, it is not constructed to model individual social media users' attitudes and motivations. As discussed previously in the digital divide literature (van Dijk, 2005), such attitudinal and motivational factors are key to modeling more complex causal and sequential relationships between technology access and their purposeful usage that is moderated by demographic, economic, and other factors. Such modeling – hopefully possible with reliable data will shed more light on underlying mechanisms that influence individuals to use social media for a diverse array of purposes ranging from social networking, as a medium of self-expression, as a forum for communal organization, to influencing, branding, and self-promotion. This is outlined as a future research direction.

It would also be important to examine how social media access and usage differs between metropolitan, micropolitan, and rural parts of the United States, in light of increasing evidence that social networking, for example using Facebook differs very negligibly between urban and rural users. However, broader use of social media, for example for content creation and sharing using platforms such as YouTube and Instagram is much more limited among rural areas compared to their urban counterparts. Examining such differences and analyzing their underlying mechanisms is essential for the creation of a more free, equitable, and inclusive digital society in the United States.

REFERENCES

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93-115.

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.

Auxier, B., and Anderson, M. (2021). Social Media Use in 2021. Pew Research Center. Retrieved from https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

Blank. G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35(6), 679-697.

Campos-Castillo, C., and Laestadius, L.I. (2020). Racial and ethnic digital divides in posting COVID-19 content on social media among US adults: secondary survey analysis. *Journal of Medical Internet Research*, 22(7):1-14.

Chen, W. (2013). The implications of social capital for the digital divides in America. *The Information Society*, 29(1), 13-25.

Esri. (2020). Methodology Statement: 2020 Esri Market Potential. Retrieved from https://downloads.esri.com/esri_content_doc/dbl/us/J9672_Market_Potential_DB_Methodology_Statement_2020.pdf.

Grubesic, T. (2006). A spatial taxonomy of broadband regions in the United States. Information and Economic Policy (18), 423-448.

Kaufmann, M., Siegfried, P., Huck, L., and Stettler, J. (2019). Analysis of tourism hotspot behaviour based on geolocated travel blog data: the case of Qyer. *International Journal of Geo-Information*, 8, 493; doi:10.3390/ijgi8110493.

Kvasny, L., and Keil, M. (2006). The challenges of redressing the digital divide: a tale of two U.S. Cities. *Information Systems Journal* (16:1), 22-53.

Lantz, B. 2013. “The large sample size fallacy,” *Scandinavian Journal of Caring Sciences* (27): pp. 487-492.

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.

Martin, M.J.R., Matthews, S.A., and Lee, B.A. (2016). The spatial diffusion of racial and ethnic diversity across U.S. counties. *Spatial Demography*. 5:145-169.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* (37:1), pp.17-23.

Nishida, T., Pick, J. B., & Sarkar, A. (2014). Japan′ s prefectural digital divide: A multivariate and spatial analysis. *Telecommunications Policy*, 38(11), 992-1010.

NTIA. 2019. *Digital Nation Data Explorer*. Washington, D.C.: National Telecommunications and Information Administration, U.S. Department of Commerce.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Geobooks, Norwich, England.

Pew Research Center. (2021). Social media fact sheet. April 7, 2021. Washington, DC: Pew Research Center.

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.

Pick, J., Sarkar, A., & Rosales, J. (2019). Social media use in American counties: geography and determinants. *ISPRS International Journal of Geo-Information*, 8(9), 424-449.

Prieger, J.E. 2013. The broadband digital divide and the economic benefits of mobile broadband for rural areas. *Telecommunications Policy* (37):483-502.

Rathore, A.K., Kar, A.K., and Llavarasan, P.V. (2017). Social media analytics: literature review and directions for future research. *Decision Analysis* 14(4):229-249.

Smith, A., Anderson, M., and Caiazza, T. (2018). Social media use in 2018. Washington, DC: Pew Research Center.

Statista. (2021). Distribution of Facebook users in the United States as of April 2021, by age group. Statista Inc. Retrieved from https://www.statista.com/statistics/187549/facebook-distribution-of-users-age-group-use.

Sullivan, G. M. and Feinn, R. 2012. “Using effect size – or why the P value is not enough”, *Journal of Graduate Medical Education* (September), pp. 279–282.

U.S. Senate. (2018). The Geography of Social Capital in America. SCP REPORT NO. 1-18, A report of the Joint Economic Committee, Retrieved from https://www.jec.senate.gov/public/_cache/files/e86f09f7-522a-469a-aa89-1e6d7c75628c/1-18-geography-of-social-capital.pdf.

Van Dijk, J. A. (2005). *The deepening divide: Inequality in the information society*. Sage Publications.

van Dijk, J.A. (2013). A theory of the digital divide, in Ragnedda, M., and Muschert, G.W. (eds.), *The Digital Divide: The internet and social inequality in international perspective*, pp. 29-51. London, England: Routledge.

Van Deursen, A. J., & Van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New media & Society*, 16(3), 507-526.

Vincente, M.R., and Lopez, A.J. (2011). Assessing the regional divide across the European Union-27. *Telecommunications Policy* (35):220-237.

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.

Yang, J., Yu, M., Qin, M., Lu, M., and Yang, C. (2019). A Twitter data credibility framework – Hurricane Harvey as a use case. *International Journal of Geo-Information*, 8, 111; doi:10.3390/ijgi8030111.

DECISION SCIENCES INSTITUTE

Understanding AI Adoption from the Lens of Trust: the Case of AI-powered DSS

Tianling Xie
University of Toledo
Email: tianling.xie@rockets.utoledo.edu

Marcelo Alvarado-Vargas
University of Toledo
Email: marcelo.alvaradovargas@utoledo.edu

ABSTRACT

As Artificial Intelligence (AI) technology advances, more AI-powered decision support systems (DSS) start to be provided by start-up AI companies and existing DSS vendors. Meanwhile, doubts and distrusts against AI may prohibit firms from adopting these systems because of concerns like AI replacing humans and the explainability issue. However, little literature has examined the role of trust in AI adoption at the corporate level. Therefore, we proposed a trust-centered adoption model for AI-powered DSS, which not only explains AI adoption from a new angle, but also provides practical implications for building trust in AI.

KEYWORDS: Artificial Intelligence; DSS; Trust; Technology Adoption; TOE Framework

INTRODUCTION

Artificial Intelligence (AI), known as the technology that simulates human behaviors, is increasingly being regarded as the trend of the future. The core technology of AI is machine learning (ML), which is a type of algorithms that can learn from the data by themselves. A great variety of business applications are based on ML and other ML-supported technologies like natural language processing and computer vision. Examples of these applications include chatbots, robots, fraud detection, and resume screening. These applications range over different business function areas. We refer to these specific applications as “Narrow AI,” in contrast with the “Artificial General Intelligence” portrayed in movies and televisions as a completely human-like intelligent agent. Multiple recent survey results indicated that managers felt compelled to adopt AI to catch the trend: Deloitte’s report (Hupfer, 2020) shows 37% of the adoption rate, and 50% of respondents in the McKinsey Global Survey (McKinsey, 2020) believed they would adopt AI in at least one business functions.

These managers’ decisions were based on valid reasons. In recent years, AI has been reported in the news, company press, and technology conferences to have the capacity to provide tremendous benefits to business operations. Its superior predictive power, for instance, is one of the most mentioned advantages, which can be applied to forecasting customer demand (van Nguyen et al., 2020), product returns (Cui et al., 2020), and other business data. Another merit of AI is that it provides operation efficiency by automation (Dash et al., 2019). Therefore, we also witness the burst of AI vendors, especially in decision-making systems (DSS). Some of them are provided by start-up companies, such as Afresh, a supply chain DSS adopted by Fresh Thymes and Heinen’s (Afresh, 2021); others are existing reputable DSS vendors: they are making efforts to integrate AI into their existing services. For example, SAP has integrated conversational AI agents and ML predictive modeling into their service products (SAP, 2021);

Oracle also has two applications powered by AI: supplier categorization and intelligence payment discount (Oracle, 2021).

Despite AI's benefits and the availability of vendors, doubts and distrusts about AI is still a common inhibitor for AI Adoption. For instance, it is found that many patients are resistant to go to healthcare providers with AI services(Longoni et al., 2019). Another example is an AI trust report by KMPG this year(KPMG Australia, 2021), reporting that most citizens surveyed from US, Germany, Canada, UK, and Australia were unwilling or ambivalent about trusting AI in healthcare and human resource systems. Partially it's originated from the technology itself: Artificial Neural Network, which is one of the most popular ML algorithms, is often considered as a black box. The users of the program can only see the data fed into the program and the predictive outcome of it, without too much understanding of the process in between. Explainability, therefore, becomes the next urgent problem to solve for researchers and practitioners. On the other hand, numerous movies, TV shows, and novels have depicted the potential adverse effects of using AI: AI may replace human labor and eventually harm human beings. People may hesitate to use AI-powered products with the impact of the media. Other issues that have raised doubts as well, including AI-related ethical problems: for instance, who's accountable for the decisions made by AI? What type of data shall be obtained and analyzed by AI without intruding on customers' privacy?

Therefore, trust has been given importance to AI adoption. But as far as we can search, limited AI adoption literature at the corporate level has considered trust a key construct. Therefore, in this study, we proposed a conceptual model for AI adoption following the Technology-Organization-Environment (TOE) framework and Diffusion of Innovation (DOI) theory, with the trust construct and the antecedents of trust. This research contributes to the literature by revealing the role of trust as a corporate psychological factor in technology adoption decisions. It also offers suggestions on building trust from multiple aspects of our society. The rest of the article will proceed as follows: first, we will examine the relevant literature; second, we will describe our conceptual model and justify the constructs and relationships in the model by theories and literature; at last, we will discuss how to apply this model and the implications for managers and future researchers.

LITERATURE REVIEW

Artificial Intelligence adoption

There have been several researches studying AI adoption using different methods and from different countries and industries. Table 1 summarized previous AI adoption researches at the corporate level. Since AI is an emerging technology, qualitative methods were adopted at the early stage to identify antecedents of AI adoption. Rao (2017) interviewed 11 South African managers following the TOE framework, and found out that technical challenges, lack of management support and organizational innovativeness, and formalization (government policy, workforce impact, etc.) inhibited AI adoption, while environmental factors such as mimetic and normative power also impacted adoption. Kessler & Schlög et al. (2019) interviewed 19 employees from various industries to explore the AI's areas of use and the drivers and challenges of adopting AI. Their respondents were unaware of AI applications' potential in many business areas, and they were primarily challenged by employee resistance and data privacy and security issues. On the other hand, Blomberg & Moberg (2019) interviewed five employees within organizations who already had AI adoption intentions for different areas, including chatbot, logistics, accounting, decision-making, and hardware development. They found top management support as a critical positive driver with the championship as a promotion tool for AI adoption, but technology competence, external support, and competitive pressure were not influential in that study.

With the same intention to explore the AI usage phenomenon, Grover et al. (2020) and Kushwaha et al. (2020) used social media analytics to discover people's attitudes and thoughts. Grover et al. (2020) were interested in AI applications in operation management. They conducted a literature review and summarized six factors for AI utilization: job fit, complexity, perceived consequences, affect toward use, social factors, and facilitating conditions. Afterward, they analyzed Twitter data using these dimensions and found that job-fit dominantly contributes to AI utilization, while trust is the highest emotion for the usage. Complexity, on the other hand, had a minor negative impact on AI utilization. Kushwaha et al. (2020) also investigated Twitter opinions about AI applications for small and medium-scale companies. Their word cloud showed that operational efficiency improvement, automation, and natural language processing were anticipated to have positive effects, while employee resistance, lack of skills, the feeling of uncertainty were the negative factors for AI usage. öhnk et al.(2021) interviewed 25 experts and summarized five AI readiness factors, including strategic alignment, resources, knowledge, culture, and data.

Survey studies about AI adoption also found similar results. Chen (2019) collected opinions from the Telecom Industry in China. He concluded that the success factors of AI adoption included compatibility, relative advantage, complexity from the technological side, managerial support from the organizational aspect, and government involvement and vendor partnership from the environment. Alsheibani (2020) surveyed Australian companies, and other than the same compatibility, relative advantage, and government regulation, their results also supported that managerial obstacles and organizational readiness were key antecedents of AI adoption. Managerial obstacles in the study referred to the challenges faced by managers, and organizational readiness included the availability of data and AI talents. Zerfass (2020) studied European companies' perspective for using AI in communication management and concluded that the lack of understanding and competence made these firms struggle to adopt AI. Pillai & Sivathanu (2020) surveyed Indian human resource (HR) and Talent Acquisition (TA) managers about their perceptions of using AI in TA. Their results approved most antecedents mentioned in other studies for adoption intention. They also supported the effect of task-technology fitness and stickiness to traditional methods.

To summarize, previous literature mainly followed the TOE framework, and the common drivers are top management support, relative advantage, government support, and the shared inhibitors are lack of resource, capabilities, and understanding related to AI. However, except for Grover et al. (2020) 's study, few research articles at the firm level included trust in their adoption framework

Table 1 Prior corporate-level research on AI adoption

Methodology	**Literature**	**Theory**	**Supported Adoption Antecedents**	**Target Population**	**Industry**
Interview	(Schlög et al., 2019)	TOE	Human resistance, security and privacy, data sufficiency	European companies	Multiple industries
Interview	(Blomberg & Moberg, 2019)	TOE	Data availability, technology competence, benefits/barriers, championship, management support, organization size, competitive pressure, external support	Organizations had decided to adopt at least one AI-capability	Multiple industries
Interview	(Rao, 2017)	TOE, DOI, Institutional Theory	Technology (integration, complexity, readiness), Organization (expected benefit, top management support, organizational innovativeness, formalization), Environment (mimetic, normative)	South African organizations	Multiple industries
Interview	(Jöhnk et al., 2021)	N/A	Strategic alignment, resources, knowledge, culture, data	AI experts	N/A
Case Study	(Lacity et al., 2021)	Action Principle Approach	N/A	Robotic and automation process adopters	Multiple Industries
Survey	(Chen, 2019)	TOE, DOI	Technology (Compatibility, relative advantage, complexity), Organization (managerial support, managerial capability), Environment (government involvement, vendor partnership)	Chinese Telecom companies	Telecom
Survey	(Alsetoohy et al., 2019)	TOE	Relative advantage, compatibility, complexity, reliability, cost, human resources, information intensity, market capabilities, attitude.	Chinese Telecom companies	Hotel food supply chain management
Survey	(Alsheibani et al., 2020)	TOE, DOI	Technology (Relative advantage, compatibility), Organization (top management support, organizational	N/A	Multiple industries

			readiness, managerial obstacles), Environment (government regulation).		
Survey	(Zerfass et al., 2020)	TOE	Organization (formal and informal linkages, communication processes) Technology (technology support, infrastructure, government regulation)	European companies	Communication management
Survey	(Pillai & Sivathanu, 2020)	TOE, TTF	Technology (cost effectiveness, relative advantage, security and privacy), Organization (HR readiness, Top management support), Environment (competitive pressure, Vendor support)	HR and TA managers from IT/ITeS companies in India	Human Resource
Literature Review & Social Media Analytics	(Grover et al., 2020)	N/A	Job-fit, complexity, perceived consequences, affect toward use (i.e., trust), social factors, facilitating conditions	N/A	Operation Management
Social Media Analytics	(Kushwaha & Kar, 2020)	Barrier Theories	Operational efficiency, unsurety and unclearness, negative employee experience.	N/A	N/A

* TOE: Technology–organization–environment (TOE) framework * TAM: Technology acceptance model *TTF: Task-Technology Fit *TPB: Theory of planned behavior *DOI: Diffusion of Innovation Theory

Artificial Intelligence and trust

Though trust has been hardly examined empirically in corporate-level AI adoption studies, it has been considered a significant factor in the many discussions about AI. Ryan (2020) explained to the reader from the trust theory that, because AI cannot be held accountable for their actions, only calculative side type of trust can be established, while the emotional aspects of trust, such as normative trust and affective trust, are not relevant to AI. Glikson (2020) summarized from the literature review four types of AI characteristics that can contribute to trust: tangibility, transparency, reliability, and immediacy behaviors. They listed the manifestations of these characteristics in three formats of AI: robot AI, virtual AI, and embedded AI. For AI in decision support systems, which is embedded AI, transparency about how the AI algorithms work will help build trust; trust is higher in tasks requiring low human intelligence; behavior-tracing, which may be used for AI analysis, decreases human trust in AI. Ferrario et al. (2020) stated that the trust between AI and humans is built incrementally, starting from the simple pragmatic belief that AI can improve the business, to epistemic reasons for reflective trust: rational understanding of how AI works, what benefit can be achieved from using certain AI functions. Rossi (2019) provided a few practical implantations for building trust, including explainability, bias mitigation, producers' behaviors to increase trust. Toreini et al. (2020) built a framework of trustworthy machine learning technologies, featuring fairness, explainability, auditability, and safety as four trust-enhancing chrematistics.

Trust has also been examined empirically to study the acceptance of other forms of AI technology at the individual level. The attitude to autonomous driving, for instance, is often considered to be impacted by trust (Panagiotopoulos & Dimitrakopoulos, 2018; Ribeiro et al., 2021; Shi et al., 2021). Social elements like social presence (De Cicco et al., 2020; Park et al., 2021) were believed to contribute to trust in is the usage of conversational agents (CV) or service robots because of the high resemblance between these AI technologies and human beings. Other antecedents of trust have also been explored, including personality traits (Bawack et al., 2021) and privacy concerns (Park et al., 2021). In these empirical studies, trust had a positive impact on users' attitude to AI directly, or indirectly through the perceived usefulness, perceived ease of use, or performance expectancy.

Therefore, based on our literature review, we were motivated to mend the gap and examine the role of trust in AI-powered decision-making systems adoption. Our theoretical model was built on the foundation of the TOE framework, DOI theory, and previous literature about trust in Information System research. More details will be described in the next section.

Trust in Information systems literature

Trust plays a vital role in IS research about attitude, behavior, and satisfaction toward technologies. Since trust has multiple layers of meanings, it has been studied in previous technology adoption literature from multiple angles. Al-Natour et al.(2008) empirically proved that perceived decision similarity between the user and the technology positively impacted the trust in the decision-aid technologies, and trust also positively contributed to the re-use intention. These authors applied the same logic to a new type of decision-aid technology: the online shopping assistant. They proved that both the perceived- decision similarity and the perceived personality similarity increased users' trust (Al-Natour et al., 2011).

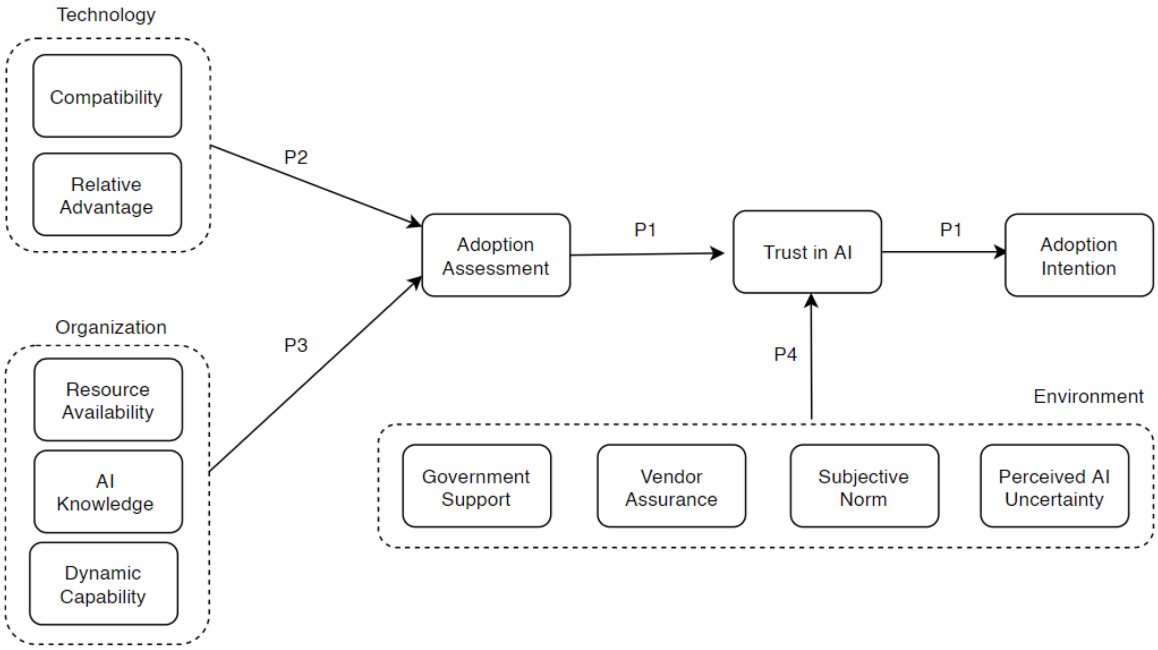
Melorose et al.(2003) identified the calculative-based, institution-based structural assurance, institution-based situational normality, and knowledge-based familiarity as trust drivers for online shopping. Calculative-based trust is about the analysis of costs and benefits; Institutional structure assurance comes from social structures, such as legislation, guarantees, and third-party assurances; The situational normality captures the reasons for trust by how normal the situation of benefiting from the technology is. Knowledge-based familiarity means the amount of information about the technology. Wang & Benbasat (2008) expanded this classification and summarized the reasons to trust recommendation agents into six types: dispositional, institutional, heuristic, calculative, interactive, and knowledge-based. Dispositional reasons are individuals' general attitudes to the technology before any interaction; Heuristic reasons are related to the look and feel of the technology; interactive reasons are based on the interactive experience with the technology. In their study about recommendation agents, dispositional reasons, calculative reasons, interactive reasons, and knowledge-based reasons were significant determinants of trust.

One stream of research divided trust into the rational and irrational aspects, which is also known as the cognitive and affective (or emotional) trust. For example, in recommendation agent studies, Komiak & Benbasat (2006) believed that the perceived personality and familiarity increased cognitive trust, and cognitive trust increased emotional trust and eventually contributed to the adoption of using recommendation agents as decision aids. Wang et al.(2016) 's study on recommendation agents believed that the presence of explanation facilities contributed to the cognitive trust, while the presence of avatars increased the affective trust. Similarly, in AI studies, Shi et al.(2021) proved in their study about autonomous driving acceptance that performance expectancy and perceived personalization contributed to the cognitive trust, while anthropomorphism and social influences impacted emotional trust. Saffarizadeh et al. (2017) stated in their conceptual model for users' information disclosure to conversational agents that, privacy concerns contributed to both cognitive and emotional trust. Some researchers focused on the rational side of trust and examined the effects of perceived risks and perceived benefits. Examples include Tan & Thoen's (2002) trust study for electronic commerce and Li et al. 's (2014) study for the adoption of personal health record systems.

THEORETICAL MODEL

The technology–organization–environment framework (TOE) has been used in numerous technology adoption studies. Tornatzky & Fleischer (1990) initially raised in their book that the factors impacting technology adoption shall not only come from technological aspects, but also from organizational and environmental aspects. As mentioned in previous sections, almost all AI adoption research followed the TOE framework to construct their survey or interview questions. To build on previous literature, we also selected our constructs and grouped them into the TOE categories in our research model (Figure 1).

Figure 1 Conceptual Model



Trust

Previous TOE-based AI adoption research treated all factors equally as exogenous factors. However, this model structure does not help us understand why certain companies with adequate resources, capabilities, and needs for AI still choose not to adopt AI. We believe that the lack of trust is a significant barrier to AI adoption. Therefore, in this study, we separated the adoption intention into two constructs, adoption potential and trust, and then proposed the model in Figure 1. This model explains the relationship among the adoption assessment, trust, and adoption intention, as well as the factors contributing to them.

The classic definition of trust by Mayer et al. (1995) is “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustors.” In the case of trust in AI-powered DSS for this study, we define trust here as “the willingness to use the information and suggestions from AI algorithms- powered systems to assist the corporate decisions process.” Trust is a necessary element for AI-based DSS adoption because managers need to believe that the suggestions of the DSS will benefit the performance of their companies to use it. We can draw a parallel between the AI-powered DSS and a weight scale. No one will use a weight scale if he or she does not trust that the weight displayed on it is accurate. In the same way, if managers don’t trust that the predicted results or decision suggestions by AI algorithms are accurate and helpful, they will not adopt them.

We believe that trust in AI can be evaluated collectively at a corporate level to study companies’ adoption intentions, since it’s a corporate belief impacting the corporate action. Several previous IS studies have also studied trust for corporate-level technological decisions. For instance, Yu et al. (2018) examined the mediating role of institutional-based trust and IT artifact trust for small and medium enterprises' cloud service transformation intention. Nicolaou et al. (2013) also used competence trust and goodwill trust as mediators between information quality and intention to use the inter-organizational electronic data exchange systems. Therefore, we propose the following:

Proposition 1: Trust serves as a mediator between adoption assessment and adoption intention.

Since the use of corporate-level DSS involves less of the emotional incentives, we mainly selected the rational side of the trust antecedents from previous trust literature, including calculative-based beliefs, institution-based structural assurance, knowledge familiarity (Gefen et al., 2003; Wang & Benbasat, 2008), and situational normality (Gefen et al., 2003). We used the adoption assessment construct to capture the calculative beliefs and knowledge-based reasons, which were included in the technological and organizational factors, while structural assurance and situational normality come from the external environment.

Adoption assessment

We define "adoption assessment" as the degree to which a firm is assessed to have the potential to adopt a technology. This construct can be viewed as an objective assessment of the possibility that a firm adopts certain technology based on its resources, capabilities, and needs. Most adoption determinants based on the TEO framework and DOI theory essentially contribute to adoption assessment. The unique contribution of this article is to separate the objective assessment from subjective mentalities. Therefore, we used the construct "adoption assessment" to capture the calculative cost-and-benefit believes related to AI adoption. Though the TOE framework gives us three broad aspects, it did not specify the concrete elements for research. With the case of AI-powered DSS, we selected the two most relevant technological factors and two organizational factors contributing to adoption assessment, based on previous AI Adoption literature and the context of AI decision-support systems.

Technological factors

Diffusion of innovations theory (DOI) explains how innovation spreads over a population or a system (Rogers, 1983). Five major factors influencing this spreading process are relative advantage, compatibility, complexity, trialability, and observability. Here we believe relative advantage and compatibility are the two major technological factors for AI adoption.

<u>Relative advantage</u>
Relative advantage means the new technology's perceived superiority compared to other available alternatives (Rogers, 1983). Companies must involve the comparing behavior to assess an emerging technology. We believe relative advantage is the primary reason why companies are considering adopting AI. Compared to existing DSS, AI has superiority regarding its higher predicting precision, more efficiency brought by automation, and more intuitive operation if more visualization is involved in the DSS. However, AI may have disadvantages concerning the financial cost. Building infrastructure, collecting data, and providing employee training for using AI involves extra costs. Some researchers use the construct "expected benefit" or "net benefit" in their model, but we believe that the net benefit associated with it may also be part of the relative advantage since a comparing process is involved. We suggest operationalizing the relative advantage from multiple dimensions, such as relative operational advantage and relative financial advantage. Relative advantage is also the most common factor

in previous AI adoption research (Alsetoohy et al., 2019; Alsheibani et al., 2020; Chen, 2019; Pillai & Sivathanu, 2020).

Compatibility

Compatibility measures the degree to which the new technology fits with the existing technology, infrastructure, resources, believes, experience, and culture (Rogers, 1983). Compatibility is of great relevance because AI differed significantly on how it works, how it should be used, the skills, talents, and data it required from traditional analytical systems. There may involve a considerable culture shock to the employees as well. Adopting AI often means buying new technology, updating the current IT infrastructure, hiring AI talents, and investing more into training. Also, as we mentioned before, Al-Natour et al.(2008) 's study proved that users need to see a similar decision-making process in the decision-aid technology to trust it, which is a form of combability with users' previous decision-making experience. This study explains why the lack of explainability is a significant barrier to AI adoption: managers' previous decision-making experience is based on rational analysis instead of blindly taking the advice from the AI black box. Therefore, we believe that compatibility issues will possibly inhibit the adoption of AI-based DSS. Compatibility is also a commonly shared construct in existing AI adoption research (Alsheibani et al., 2020; Chen, 2019; Rao, 2017).

Other factors in DOI theory, namely, complexity, trialability, and observability, were not included in the model. First, the AI-powered DSS is not necessarily more complex compared to traditional DSS tools. For instance, SAP Analytics Cloud applies conversational analytics to their information query system so that users can type in questions in a manner of natural conversations to obtain analytics reports. And most software provides demos and trials and therefore we do not consider trialability as a significant technological factor that makes a difference in the adoption assessment. We also decided not to include observability as a technological factor; observability is defined as the degree in which the innovation provides tangible results by DOI, but it is largely determined by the information received from the external environment, such as news, industry partners, and the government. We will cover these elements in the environmental factors. Besides, relative advantages and compatibility are also the two constructs that are mostly proven in previous AI adoption studies.

Proposition 2: Technological factors impact a firm's adoption assessment for AI-power DSS.

Organizational factors

Resource availability

Resource Availability is defined in this study as the availability or sufficiency of financial resources, human resources, and technological resources. Needless to say, all investments in technology innovation need to be financially supported. Previous literature often considered "financial healthiness" as a predictor of the financial outcome of technology adoption (David et al., 2001; Foster & Levitan, 2006). Human resources, including existing IT experts, and potentially hired AI experts, will also impact the decision to adopt AI-DSS. As mentioned in the previous section, lack of AI talents is one of the managerial obstacles in Alsheibani's (2020) AI adoption survey. öhnk et al.(2021) also listed "resources" as one of the AI-readiness factors, including the financial budget, personnel, and IT infrastructure. As AI is an emerging technology, it is unlikely that non-tech companies will have existing experts who understand how to deploy an AI system. Nevertheless, the available knowledge of knowing where to gain these talents exists in the competent human resource management teams, and the knowledge of how and where to gain support and training resides in existing IT teams.

AI knowledge
As suggested by Kessler & Schlög et al. (2019) and Zerfas's (2020) research, many managers and employees are unaware of the potential of AI applications and how to use them. Without relevant knowledge, it would be difficult for managers to develop a strategy to embrace the new technology. Managers without appropriate AI knowledge would also be less likely to appreciate the opportunities brought by the technology or to assess the cost and benefit associated with AI. Therefore, öhnk et al.(2021) identified "knowledge" as one of the AI readiness factors, including AI awareness, upskills, and AI ethics. Furthermore, the fear of AI replacing human jobs also may come from a lack of knowledge. Therefore, we believe the overall knowledge level about AI will contribute to how it will assess the potential adoption of AI-power DSS. Also, because knowledge-based beliefs have been empirically proved to increase trust in technologies (Gefen et al., 2003; Wang & Benbasat, 2008), we also believe AI knowledge will increase the trust in AI through the adoption assessment.

Dynamic capability
Dynamic capability, defined as "the firm's ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments," (Teece et al., 1997) is considered an important factor for organizations to cope with disruptive technologies in the organizational research literature. Dynamic capability determines how fast a company can sense a need or an opportunity for change in the external environment, reorganize and reconfigure the internal resources a firm has, and then implement the change strategically. In information system literature, dynamic capability included as a driving force for adopting technology-based innovations that requires transformative changes. For instance, Soluk & Kammerlander (2021) described details of how dynamic capabilities are required in each step of the digital transformation process through 127 interviews of family-owned Mittelstand firms. Mikalef et al. (2020) consider dynamic capabilities as a mediating role between a firm's big data analytics capabilities and its marketing and technological capabilities. We believe dynamic capability is a critical organizational factor influencing a firm's assessment for potentially adopting AI-powered DSS, because when AI is involved, there will be a higher degree of automation in decision-making, as well as a higher requirement of data collection and processing. Therefore, the pre-adoption configuration of business processes and resources need to be updated to match the new requirements, and this change will require the dynamic capability.

We did not include top management support as an organizational factor because for corporate-wide adoption decision for an innovative information technology like AI, management support is a determining factor, and all other factors such as relative advantages and resource availability are the factors driving the management support. And implementations of such technologies are usually in a top-down manner with management support. Therefore, management support can be considered as equal to adoption decision and therefore, it is not very meaningful to examine this factor's effect on adoption decision.

Proposition 3: Organizational factors impact a firm's adoption assessment for AI-power DSS.

Environmental factors

Government support

Government support includes the policies, rules, and regulations related to AI. A sense of uncertainty was mentioned by previous AI adoption literature as an inhibitor (Kushwaha & Kar, 2020). Government support implies that the usage of AI is formalized and normalized, thus reducing uncertainty and increasing trust. The support from the government can be financial support, such as grants and funds, or the existing regulation and policies. Also, government support or government regulation was identified as an antecedent in most previous literature (Alsheibani et al., 2020; Chen, 2019; Rao, 2017). Government support can also be viewed as an institutional structure assurance reason to trust AI (Wang & Benbasat, 2008).

Vender assurance
Vender assurance is another form of institutional assurance for trust (Wang & Benbasat, 2008). We define it as the perceived possibility of success for using AI-powered DSS from vendors' marketing materials and technical support. Many AI vendors have listed their success cases and stories to give customers confidence that their AI product would effectively provide the guaranteed benefits. Also, many reputable vendors have already assumed the responsibility to take the lead in building trust. For instance, IBM has a segment under their AI research related explicitly to trusted AI. Other big firms such as Google have given their principles on AI development, ensuring their privacy protection, fairness, and transparency to increase trust.

Subjective norm
Subjective norm is the perceived social pressure to perform or not to perform certain behavior (Ajzen, 1991). When the powerful entities in the environment the firm was situated in have adopted AI-DSS, or when most of the companies that impact the firm have adopted AI, the managers and employees of the firm may also believe that it is necessary to adopt AI-DSS as well. Examples of these firms could be the industry leader, competitors, customers, industry partners, and suppliers. Subjective norm can be viewed as a close concept to situational normality, which means the perceived normality to use the new technology. Situational normality is an important source of trust in previous literature (Gefen et al., 2003).

Perceived AI uncertainty
Song (2001) defined the perceived technology uncertainty as "the inability to completely understand or accurately predict some aspect of the technological environment." Kushwaha & Kar (2020) mentioned that "a sense of uncertainty" was one of the inhibitors to adopting AI. Perceived uncertainty about AI comes from the external environment where both positive and negative opinions about AI coexists. For instance, we often hear news about exciting innovations accelerated by AI, such as drug discovery or AI robots, but in the meantime, negative information about AI also cannot be ignored, such as bias issues related to AI hiring. These different voices in the external environment create a sense of confusion that makes managers unsure about the expected benefit of using AI. The portrait from media that AI may eventually replace human and take over the earth also worsen the uncertainty; people may fear the increased development of AI could lead to the future disasters. We can compare AI with other technologies at their early stage, such as the Internet. Internet has become a necessity in the current day and age; however, at an early stage of the Internet, articles like "IT doesn't Matter" represented an option that firms shall not rush into adopting Internet as it had not reached its maturity stage and did not worth taking the risk to invest in it. Concluding the four environmental factors mentioned above, we proposed the following:

Proposition 4: Environmental factors impact a firm's Trust in AI.

Table 2 summarized all the constructs and their definitions:

Table 2 Summary of Constructs and Definitions

Constructs	Definitions
Adoption assessment	The degree to which a firm is assessed to have the potential to adopt a technology.
Trust	The willingness to use the information and suggestions from AI algorithms-powered systems to assist the corporate decisions process.
Technological Factors	
Relative advantage	The perceived superiority of AI-powered DSS compared to other available alternatives.
Compatibility	The degree to which the new technology fits with the existing technology, infrastructure, resources, believes, experience and culture.
Organizational Factors	
Resource availability	The availability or sufficiency of financial resources, human resources, and technological resources
AI knowledge	The firms' existing understanding of Artificial Intelligence and its application on the decision support systems.
Dynamic capability	The firm's ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments
Environmental Factors	
Government support	Government support includes the policies, rules, and regulations related to AI.
Vender assurance	The perceived possibility of success for using AI-powered DSS from vendors' marketing materials and technical support.
Subjective norm	The perceived social pressure to adopt AI-powered DSS.
Perceived AI uncertainty	The inability to completely understand or accurately predict some aspect of the AI technological environment.

DISCUSSION AND CONCLUSIONS

In this article, we proposed a trust-centered theoretical model for AI adoption. Our model differed from how previous literature formulated direct relationships from technological, organizational, and environmental factors to AI adoption intention. Instead, we separated the objective AI adoption assessment from the subjective trust and proposed that environmental factors impact the trust construct instead of the adoption intention directly. The mediating role of the trust element explains why certain companies with sufficient financial and technological capabilities still refrained from jumping into the rush of adopting AI. We also explained in our model where the sources of trust and distrust may be. The theory of trust also provides a new lens to look at the traditional views of diffusion of innovations. DOI classifies the innovation adoption population into five categories: innovators, early adopters, early majority, late majority, and laggards. From the lens of trust, companies adopt new technologies at different timing because they have different levels of trust in the benefits that this new technology may bring. Therefore, to speed up the diffusion of technology within an industry or a country, other than improving the technology itself, we can increase the trust level from the industry or country's overall perspective.

The same framework can be applied to other emerging technologies, such as mobile payment, 5G network, and blockchain. Distrusting elements exist for these technologies that prevent the diffusion of them, such as the conspiracy surrounding 5G network, and the security concerns for mobile payment. Even though these technologies may have relative advantages over existing alternatives, lack of trust prevents them from being widely adopted in certain countries or industries.

This research may benefit companies considering AI-powered DSS superior in its abilities but have not developed a plan for adopting it. By looking at the model, they may assess their readiness to adopt AI by asking themselves: can we solve compatibilities issues? Do we have sufficient financial, technical, and human resources for adopting AI? Do we know how to use AI to reach its best potential? For those companies distrusting AI, it may be beneficial for them to use our research as a mind map and re-assess where their distrust comes from and help them make better adoption decisions.

Our research also has practical implications for AI vendors to achieve better product acceptance. To solve the compatibility issues, AI vendors provide evidence on how their products can be integrated into customer firms' existing operations in a short time, or to offer customers best practices of operational adjustment. Therefore, the function of vendor support shall extend from technical support only to operation integration as well. It may also be beneficial to foster communities and networks of AI talents that can help with learning and hiring. Existing DSS vendors who intend to integrate AI into their existing DSS products shall also investigate customer companies' practices to make a better strategic adjustment on their DSS products. Pricing strategies need to be carefully considered as financial resources could be an important constraint for adoption. Since some firms may be unaware of what AI-powered DSS is and what it can do, it is a suggested strategy to provide a demo or trial of the AI solution so that their AI knowledge may increase after using it. Moreover, webinars and training may increase the AI knowledge familiarity and increase trust and adoption for the AI products. AI vendors may market to the industry leaders to increase the subjective norm of that industry. They may also increase their positive media exposure to decrease the perceived AI uncertainty.

This theoretical model also has general implications for building trust in AI at a societal level. The government has the responsibility to foster an affirming environment by providing regulations and policies. Leading firms may build alliances surrounding AI and share experiences about better facilitating it. Industry standards evaluating the security, fairness, and accountability of AI may help to assure the benefit of AI and reduced the perceived uncertainty. Other than what have been already mentioned, vendors can also provide case studies and third-party assurances on their AI products.

Our research is not without limitations. We only focus on a few most relevant elements in each T-O-E category. Other factors such as firm innovativeness, firm size, complexity can be considered in the future. Future researchers may also operationalize our constructs and test our propositions by methods such as survey. They could also model the diffusion process of AI mathematically and compare the results by manipulating certain trust antecedents.

REFERENCES

Afresh. (2021). *Afresh - Home*. https://www.afresh.com/

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211.

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2011). The adoption of online shopping assistants: Perceived similarity as an antecedent to evaluative beliefs. *Journal of the Association for Information Systems*, *12*(5), 347–374. https://doi.org/10.17705/1jais.00267

Al-Natour, S., Benbasat, I., & Cenfetelli, R. T. (2008). The effects of process and outcome similarity on users' evaluations of decision aids. *Decision Sciences*, *39*(2), 175–211. https://doi.org/10.1111/j.1540-5915.2008.00189.x

Alsetoohy, O., Ayoun, B., Arous, S., Megahed, F., & Nabil, G. (2019). Intelligent agent technology: what affects its adoption in hotel food supply chain management? *Journal of Hospitality and Tourism Technology*, *10*(3), 317–341. https://doi.org/10.1108/JHTT-01-2018-0005

Alsheibani, S., Messom, C., & Cheung, Y. (2020). Re-thinking the Competitive Landscape of Artificial Intelligence. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, *3*, 5861–5870. https://doi.org/10.24251/hicss.2020.718

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from SEM and fuzzy set qualitative comparative analysis. *International Journal of Information Management*, *58*(January). https://doi.org/10.1016/j.ijinfomgt.2021.102309

Blomberg, E., & Moberg, F. (2019). *Artificial Intelligence Adoption – Is it more than just hype ? A qualitative study of what factors influence an organization's decision to adopt AI*.

Chen, H. (2019). *Success Factors Impacting Artificial Intelligence Adoption - Perspective From the Telecom Industry in China By*.

Cui, H., Rajagopalan, S., & Ward, A. R. (2020). Predicting product return volume using machine learning methods. *European Journal of Operational Research*, *281*(3), 612–627. https://doi.org/10.1016/j.ejor.2019.05.046

Dash, R., McMurtrey, M., Rebman, C., & Kar, U. K. (2019). Application of Artificial Intelligence in Automation of Supply Chain Management. *Journal of Strategic Innovation and Sustainability*, *14*(3), 43–53. https://doi.org/10.33423/jsis.v14i3.2105

David, H., James, H., & Jacqueline, R. (2001). Market Reactions to ERP Implementation Announcements.pdf. In *Journal of Information Systems* (Vol. 15, Issue 1, pp. 3–18). http://www.google.ca/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0CDAQFjAB&url=http%3A%2F%2Fnubs.nju.edu.cn%2Fdownload.php%3Ffilepath%3Dupload%2Fteacherp%2Fzcd%2FMarket%2520Reactions%2520to%2520ERP%2520Implementation%2520Announcements.pdf&ei=S1NtT8zxFKPi

De Cicco, R., e Silva, S. C., & Alparone, F. R. (2020). Millennials' attitude toward chatbots: an experimental study in a social relationship perspective. *International Journal of Retail and Distribution Management*, *48*(11), 1213–1233. https://doi.org/10.1108/IJRDM-12-2019-0406

Ferrario, A., Loi, M., & Viganò, E. (2020). In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions. *Philosophy and Technology*, *33*(3), 523–539. https://doi.org/10.1007/s13347-019-00378-3

Foster, B. P., & Levitan, A. S. (2006). A Note On Market Reactions To Relatively Recent ERP Announcements. In *Review of Business Information Systems (RBIS)* (Vol. 10, Issue 3, pp. 9–18). https://doi.org/10.19030/rbis.v10i3.5332

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM In Online Shopping: An Integrated Model. *MIS Quarterly*, *27*(1), 51–90.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Grover, P., Kar, A. K., & Dwivedi, Y. K. (2020). Understanding artificial intelligence adoption in operations management: insights from the review of academic literature and social media discussions. *Annals of Operations Research*, *0123456789*. https://doi.org/10.1007/s10479-020-03683-9

Hupfer, S. (2020). *AI adoption in the workforce*. Deloitte Insights. https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/ai-adoption-in-the-workforce.html

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors. *Business and Information Systems Engineering*, *63*(1), 5–20. https://doi.org/10.1007/s12599-020-00676-7

Komiak, S. Y. X., & Benbasat, I. (2006). The Effects of Personalization Trust and Adoption of Recommendation. *Management Information Systems Quarterly*, *30*(4), 941–960.

KPMG Australia. (2021). *Trust in Artificial Intelligence: A five country study*. https://home.kpmg/au/en/home/insights/2021/03/artificial-intelligence-five-country-study.html

Kushwaha, A. K., & Kar, A. K. (2020). Micro-foundations of Artificial Intelligence Adoption in Business: Making the Shift. In *IFIP Advances in Information and Communication Technology* (Vol. 617, pp. 249–260). Springer International Publishing. https://doi.org/10.1007/978-3-030-64849-7_22

Lacity, M., Willcocks, L., & Gozman, D. (2021). Influencing information systems practice: The action principles approach applied to robotic process and cognitive automation. *Journal of Information Technology*, *36*(3), 216–240. https://doi.org/10.1177/0268396221990778

Li, H., Gupta, A., Zhang, J., & Sarathy, R. (2014). Examining the decision to use standalone personal health record systems as a trust-enabled fair social contract. *Decision Support Systems*, *57*(1), 376–386. https://doi.org/10.1016/j.dss.2012.10.043

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734.

McKinsey. (2020). *The state of AI in 2020*.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information and Management*, *57*(2). https://doi.org/10.1016/j.im.2019.05.004

Nicolaou, A. I., Ibrahim, M., & van Heck, E. (2013). Information quality, trust, and risk perceptions in electronic data exchanges. *Decision Support Systems*, *54*(2), 986–996. https://doi.org/10.1016/j.dss.2012.10.024

Oracle. (2021). *Oracle Fusion Analytics Warehouse*. https://www.oracle.com/business-analytics/fusion-analytics.html

Panagiotopoulos, I., & Dimitrakopoulos, G. (2018). An empirical investigation on consumers' intentions towards autonomous driving. *Transportation Research Part C: Emerging Technologies*, *95*(September), 773–784. https://doi.org/10.1016/j.trc.2018.08.013

Park, S. S., Tung, C. T. D., & Lee, H. (2021). The adoption of AI service robots: A comparison between credence and experience service settings. *Psychology and Marketing*, *38*(4), 691–703. https://doi.org/10.1002/mar.21468

Pillai, R., & Sivathanu, B. (2020). Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. *Benchmarking*, *27*(9), 2599–2629. https://doi.org/10.1108/BIJ-04-2020-0186

Rao, T. (2017). *Factors critical to the organisational adoption of artificial intelligence: A South African perspective* (Issue November).

Ribeiro, M. A., Gursoy, D., & Chi, O. H. (2021). Customer Acceptance of Autonomous Vehicles in Travel and Tourism. *Journal of Travel Research*. https://doi.org/10.1177/0047287521993578

Rogers, E. M. (1983). Diffusion of innovations. In *An Integrated Approach to Communication Theory and Research, Third Edition* (3rd ed.).

Rossi, F. (2019). Building Trust in Artificial Intelligence. *Journal of International Affairs*, *71*(1), 127–134.

Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics*, *26*(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y

Saffarizadeh, K., Boodraj, M., & Alashoor, T. (2017). Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-Disclosure. *Thirty Eighth International Conference on Information Systems*, 1–12.

SAP. (2021). *Artificial Intelligence | SAP Intelligent Technologies*. https://www.sap.com/products/artificial-intelligence.html

Schlög, S., Postulka, C., Bernsteiner, R., & Ploder, C. (2019). Artificial Intelligence Tool Penetration in Business: Adoption, Challenges and Fears. In L. Uden, I.-H. Ting, & J. M. Corchado (Eds.), *Knowledge Management in Organizations* (Vol. 1027, pp. 259–270). Springer International Publishing. https://doi.org/10.1007/978-3-030-21451-7

Shi, S., Gong, Y., & Gursoy, D. (2021). Antecedents of Trust and Adoption Intention toward Artificially Intelligent Recommendation Systems in Travel Planning: A Heuristic–Systematic Model. *Journal of Travel Research*, *60*(8), 1714–1734. https://doi.org/10.1177/0047287520966395

Soluk, J., & Kammerlander, N. (2021). Digital transformation in family-owned Mittelstand firms: A dynamic capabilities perspective. *European Journal of Information Systems*. https://doi.org/10.1080/0960085X.2020.1857666

Song, M. (2001). The Effect of Perceived Technological Uncertainty on Japanese New Product Development. *Academy of Management Journal*, *44*(1), 61–80.

Tan, Y. H., & Thoen, W. (2002). Formal aspects of a generic model of trust for electronic commerce. *Decision Support Systems*, *33*, 233–246.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal*, *18*, 509–533.

Toreini, E., Aitken, M., Coopamootoo, K., Elliott, K., Zelaya, C. G., & van Moorsel, A. (2020). The relationship between trust in AI and trustworthy machine learning technologies. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 272–283. https://doi.org/10.1145/3351095.3372834

Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*.

van Nguyen, T., Zhou, L., Chong, A. Y. L., Li, B., & Pu, X. (2020). Predicting customer demand for remanufactured products: A data-mining approach. *European Journal of Operational Research*, *281*(3), 543–558. https://doi.org/10.1016/j.ejor.2019.08.015

Wang, W., & Benbasat, I. (2008). Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce. *Journal of Management Information Systems*, *24*(4), 249–273. https://doi.org/10.2753/MIS0742-1222240410

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, *86*, 48–60. https://doi.org/10.1016/j.dss.2016.03.007

Yu, Y., Li, M., Li, X., Zhao, J. L., & Zhao, D. (2018). Effects of entrepreneurship and IT fashion on SMEs' transformation toward cloud service through mediation of trust. *Information and Management*, *55*(2), 245–257. https://doi.org/10.1016/j.im.2017.07.001

Zerfass, A., Hagelstein, J., & Tench, R. (2020). Artificial intelligence in communication management: a cross-national study on adoption and knowledge, impact, challenges and risks. *Journal of Communication Management*, *24*(4), 377–389. https://doi.org/10.1108/JCOM-10-2019-0137

Unravelling Institutional Factors leading to the adoption of XBRL: Context of GCC countries

Abstract:

This paper provides an investigation of the determinants of eXtensible Business Reporting Language (XBRL) adoption. This is achieved by examining a sample of 71 countries during the period from 2009 – 2014. The findings indicate that countries, with strong investor protection, higher levels of economic development, higher degrees of external economic openness, more highly educated populations and with high levels of technology absorption, are more likely to adopt XBRL. Countries, which have an Anglo-Saxon culture, however, are less likely to adopt XBRL than other countries. The burden of government regulation, the macroeconomic stability and the mandatory adoption of IFRS have no significant influence on XBRL adoption.

Keywords: XBRL; across countries; culture; investor protection; government regulation; economic development; economic openness; economic stability; education; technology absorption and IFRS.

1. Introduction

The eXtensible Business Reporting Language (XBRL) is an electronic language for commercial and financial data that revolutionizes the worldwide communication of information (Markelevich, Shaw, & Weihs, 2015; Chotaliya, 2014; Steenkamp & Nel, 2012). It was developed by XBRL International, which is a consortium of over 600 members. This consortium consists of regulatory bodies, public audit firms, financial institutions, and software vendors as well as international and national standard setters and public entities (Wang, Wen, & Seng, 2014). XBRL works like a barcode on an item in the grocery store (Manmohan & Pk, 2014). Instead of a barcode for physical products, XBRL is a barcode for information about a company (PriceWaterHouseCoopers, 2003).

Currently, numerous countries have implemented eXtensible Business Reporting Language (XBRL) to meet the information needs of securities regulators and different government agencies. Also, many pilot projects are underway in several countries (e.g., Finland and Brazil) to assess the potential benefits that XBRL can offer.

Some countries (e.g., Japan and United States of America) require the use of XBRL while others (e.g., Canada and Peru) permit its use. However, many countries (e.g., New Zealand and Tunisia) still do not mandate or permit the use of XBRL. Also, countries adopt XBRL to meet the information needs of differents users. Some countries (e.g., South Africa and Chile) adopt XBRL to meet the information needs of securities regulators. Other countries (e.g., Belgium and France) adopt XBRL to meet the information needs of other government agencies. A number of countries (e.g., India and Australia) use XBRL to meet the information needs of both securities regulators and other government agencies.

Literature on the determinants of XBRL adoption is limited at firm-level. For example, Pinsker and Li (2008) find that uncertainty, taxonomy and costs are the main factors that can affect firms' decision to adopt XBRL. Literature is rare on the determinants of XBRL adoption at the country-level. Therefore, the aim of this paper is to examine these determinants. In particular, we examine the extent to which XBRL adoption is influenced by the following factors: culture; strength of investor protection; burden of government regulation; level of economic development; degree of external economic openness; macro economic stability; education; country's firm-level technology absorption and accounting system (IFRS). Using a sample of 71 countries over the period from 2009 to 2014, we find that the countries, which are most likely to adopt XBRL, have strong investor protection, higher levels of economic development, higher degrees of external economic openness, more highly educated populations and high levels of technology absorption. Countries, which have an Anglo-Saxon culture, are less inclined to adopt XBRL. We also find that the burden of government regulation, the macroeconomic stability and the mandatory adoption of IFRS have no impact on XBRL adoption.

Our study contributes to the literature by identifying the factors that can explain why countries either adopt or do not adopt XBRL. We use an international sample and include a large number of countries. Then, we specify our model and identify the factors that could explain the country's decision to adopt or not to adopt the XBRL to meet the information needs of securities regulators and other government agencies or both. In addition, we examine the factors that motivate developed and developing countries separately to adopt XBRL. After using a pooled logistic regression, we test these models for each year in the sample period.

The remainder of this paper is organised as follows. Section 2 reviews the relevant literature. Section 3 develops our hypotheses. Data description, sample selection criteria and the empirical model are discussed in Section 4. The empirical findings are presented and discussed in Section 5. Section 6 concludes.

2. Relevant literature

Effective corporate governance implies transparency of information and adequate monitoring of information disclosure to the public (Roohani, Furusho, & Koizumi, 2009). The financial scandals have exposed corporate governance weaknesses and have attracted greater attention from investors and scholars alike about the various issues of governance and transparency. Some of the governance weaknesses stem from the information asymmetry between insiders and the investing public (Ragothaman, 2012).

Information asymmetry in the form of adverse selection and moral hazard has different adverse outcomes such as transaction costs, low liquidity market and increasing cost of capital.

In this respect, management is encouraged to reduce this asymmetry and disclose more voluntary information (Bahmani, 2014). One way to mitigate the information asymmetry problem is to improve disclosures through the adoption of XBRL (Yoon, Zo, & Ciganek, 2011). In this section, our literature review by discussing empirical studies that are concened with XBRL adoption.

XBRL tagged information is relevant to investors for investment decisions (Arnold et al., 2012; Birt, Muthusamy, & Bir, 2017). Zhang, Riordan, and Weinhardt (2013) find that analyst following increased after the voluntary introduction of XBRL. In addition, literature shows that the adoption of XBRL improves analyst forecasts accuracy (Liu, Wang, & Yao 2014; Felo, Kim, & Lim, 2018) reduces the cost of capital (Li, Lin, & Ni, 2012; Chen et al., 2015; Lai et al., 2015) and increase market liquidity (Li, Lin, & Ni, 2012; Liu, Luo, & Wang, 2017). The adoption of XBRL improves transparency, timeliness and relevance of financial reporting information (Wallace, 2001; Willis, Tesnière, & Jones, 2002; PriceWaterHouseCoopers, 2003; Wang, Wen, & Seng, 2014; Manmohan & Pk, 2014; Talebnia, 2016).

The research on the determinants of XBRL adoption is rare. Literature is limited to a specific region or context. For example, Yassin, Barghouthi, and Al-Khatib (2012) find that the main obstacles of using XBRL in Jordan are the lack of experts to implement and adopt XBRL; the time and effort required to use the XBRL language in communicating financial reporting; and the cost related to the adoptin of the system.

Gostimir (2015) argues that the main obstacle of using XBRL in Croatia is the absence of awareness among the business and regulatory communities as well as the public. Similarly, Ghani, Said, and Muhammad (2014) show that awareness is the main factor affecting the adoption of XBRL in Malasia.

In South Africa, Steenkamp and Nel (2012) find that there is no awareness about XBRL. In addition, they find that the main reasons why XBRL is not implemented by South African companies are: XBRL is not yet mandatory; there are no benefits in doing so; the necessary technical knowledge is lacking; and management has no knowledge of XBRL.

Literature also shows that the main factors responsible for the delayed acceptance of XBRL in Brazil include: the lack of public regulation; users of corporate financial information know virtually nothing about XBRL; the absence of tools for its use; and low demand from financial information users (Nobre & Carvalho, 2011).

Premuroso and Bhattacharya (2008) and Boritz and Timoshenko (2015) investigate the US's decision to adopt XBRL. These studies are limited to firm-level attributes such as firm characteristics and corporate governance. In the same context and using a survey, Pinsker's (2008) findings show that the more useful a company perceives XBRL, the greater is its intent to adopt and the higher its absorptive capacity is.

Using a sample of EU companies, Elissavet, Antonios, and Theodora (2013) develop a model that is based on the Technology Acceptance Model 2 (TAM2) in order to examine the factors that led to the acceptance and usage of XBRL. The proposed model consists of five key constructs: perceived usefulness, perceived ease of use, output quality, training costs, and behavioral intention. Their results show that XBRL's perceived usefulness has a positive effect on the user's intention to use XBRL. Users will use XBRL if they perceive its efficiency and effectiveness with respect to their jobs. Perceived ease of use has a significant effect on users' intention to use XBRL. In this regard, XBRL's training costs are not related to the intention to use XBRL. Since XBRL demands high training costs, users appear to react negatively to its use. Perceived usefulness occurs if the XBRL user perceives that XBRL is easier to use than previous existing ways of working and if the XBRL user is satisfied with the total output quality that results from the use of XBRL.

Using a qualitative study of XBRL adoption, Troshani and Doolin (2005) analyze the drivers and the inhibitors of Australia's XBRL adoption. They examine a set of factors that impact the adoption of XBRL; these are environment, organizational, and innovation context factors. They find that the local adoption strategy, the limited local XBRL success stories and champions, the priority to adopt International Accounting Standards (IAS), the employees' education, the limited resources and software tool support, and the instability of the XBRL specification are the most important factors that affect XBRL adoption.

O'Kelly (2010) indicates that the reduction of the administrative burden, the transparency of the capital market, and the adoption of IFRS are among the drivers of the adoption of XBRL.

Rawashdeh and Selamat (2013) examine the factors that may affect the adoption of XBRL among the accounting professionals in Saudi-Arabia. Their results indicate that perceived ease of use, relative advantage, compatibility, social influence, knowledge and Internet skills are the key factors driving Saudi Arabia's adoption of XBRL.

Bonsón, Cortijo, and Escobar (2009) identify the main motivation of XBRL adoption in North-American companies. These include: the intention to benefit from all advantages of XBRL; the intention to acquire a corporate image as a pioneer in technology; the ability to better use of financial information using software applications; and the intention to improve transparency.

The above-mentioned studies are based on case studies that were conducted in various contexts. Steenkamp and Nel (2012) state that the factors in the past that have driven the adoption of IFRS, are also driving the current adoption of XBRL. These factors are the investor demand for greater transparency, accuracy, accessibility and, ultimately, financial statements that are more relevant and useful for decision-making.

Several studies have investigated the factors that could explain the countries' adoption of IFRS (Hope, Jin, & Kang, 2006; Zeghal & Mhedhbi, 2006; Archambault, & Archambault, 2009; Clements, Neill, & Scott Stovall, 2010; Shima, Yang, & Boulevard, 2012; Zehri & Chouaibi, 2013; Kossentini & Othman, 2014).

Hope, Jin and Kang (2006) examine a group of institutional factors that might influence a country's decision to adopt IFRS: these were, namely, disclosure requirements; anti-director rights; and stock market access. Using a sample of 38 countries, their findings show that countries with weak shareholder protection (i.e. poor disclosure rules and anti-director rights) are more likely to adopt IFRS than countries with strong shareholder protection. Indeed, their findings show that countries, which provide better access to their stock markets for international investors, are more likely to adopt IFRS.

Zeghal and Mhedhbi (2006) investigate the factors that may affect the decision of 64 developing countries to either adopt or not adopt IFRS. The authors consider the factors of economic growth, education level, the degree of external economic openness, cultural membership in a group of countries and the existence of a capital market. Their findings show that developing countries, which enjoy the highest literacy rate, have a capital market, and that those countries belonging to an Anglo-American culture are the most motivated to adopt IFRS.

In order to understand the fundamentals behind 74 developing countries' decisions to adopt IFRS, Zehri and Chouaibi (2013) extended Zeghal and Mhedhbi's (2006) study to integration factors such as culture, economic growth, capital market availability, educationlevel, openness to the exterior world, legal system and political system. The results indicate that developing countries, which have a high level of economic growth along with a legal system of common law and an advanced level of education, are most likely to adopt IFRS.

Similarly, several studies show the factors that may influence the quality and the transparency of financial reporting at the country level (Boolaky, Krishnamurti, & Hogue, 2013; Othman & Zeghal, 2008). Boolaky, Krishnamurti, and Hogue (2013) examine the role of environmental factors that influence the strength of a country's auditing and reporting standards. The strength of auditing and reporting standards is a proxy for institutional transparency that is expected to have a major effect on the quality of the financial information produced by companies in any given country (WEF, GCR 2011-2012). They find that institutional infrastructure, financial market development, and higher education and training all jointly have a significant and positive impact on the strength of auditing and reporting standards. According to Boolaky, Krishnamurti, and Hogue (2013), good accounting and auditing regulations facilitate transparency through better disclosure of information and facilitate comparisons between companies. Strong regulations provide business incentives to provide investors with useful and relevant information. The relevant regulations include financial reporting requirements, audit standards, and generally accepted accounting principles. If regulations are weak, companies may choose to not disclose information or to manipulate the required information (Boolaky, Krishnamurti, & Hogue, 2013). As an interactive data format, XBRL is also expected to improve the transparency and the quality of financial reporting across companies and countries (Efendi, Park, & Smith, 2014; Hao, Zhang, & Fang, 2014; Shanmuganathan, 2016).

Othman and Zeghal (2008) examine country-level attributes that impact corporate governance disclosures that depend on the legal system of an emerging market country. The results indicate a significantly higher influence of law enforcement on corporate governance disclosure for common law emerging markets than for civil law ones. Indeed, in both common law and civil law emerging markets, the size of the capital market has a substantial effect on corporate governance disclosure. According to Othman and Zeghal (2008), corporate governance disclosure provides a description of companies' corporate governance characteristics. Therefore, this type of disclosure helps to convey valuable information to many decision-makers and allows them to assess the relevance and effectiveness of each company's corporate governance system. Further, corporate governance disclosure reduces the asymmetries between insiders (managers and major shareholders) and outsiders (minor shareholders). Hence, one of XBRL's goals is to reduce information asymmetry, and many studies prove the negative association between XBRL and information asymmetry (Bai, Sakaue, & Takeda, 2014; Yoon, Zo, & Ciganek, 2011; Tan & Shon, 2009).

From another perspective, the results of several studies show the factors that favor the countries' adoption of information technology diffusion (e.g., Wunnava & Leiter, 2009; Beilock & Dimitrova, 2003).

Wunnava and Leiter (2008) find that economic strength, telecommunications and technology infrastructure, proficiency in English and a country's political and economic openness play a fundamentally important role in determining diffusion rates. In addition, tertiary enrollment and income equality play an essential role in terms of Internet diffusion.

Beilock and Dimitrova (2003) is one of the earliest studies that attempts to identify and measure the effects of the major determinants of Internet diffusion at country-level. They find that per capita income is the most important determinant. The openness of a society and Infrastructure are, also, important determinants.

Wunnava and Leiter (2008) attempt to discover the main determinants concerning inter-country penetration rates of the Internet. Their results indicate that the economic strength, telecommunications and technology infrastructure, proficiency in English, and a country's political and economic openness play a fundamentally important role in determining diffusion rates. In addition, tertiary enrollment and income equality play an essential role in terms of Internet diffusion.

Beilock and Dimitrova (2003) is one of the earliest studies that attempts to identify and measure the effects of the major determinants of Internet diffusion at the country level. By using a sample of 105 developed and developing countries, they find that per capita income is the most important determinant. The openness of a society and its infrastructure are also important determinants.

This paper uses a similar approach to the above-mentioned studies. In particular, this study seeks to determine whether culture, strength of investor protection, burden of govenment regulation, level of economic development, degree of external economic openness, macroeconomic stability, education, country's firm-level technology absorption and the mandatory adoption of IFRS explain a country's decision either to adopt or not adopt XBRL. We chose these variables based on the literature review and the availability of data. We use them to develop our research hypotheses.

3. Hypotheses development

3.1.Culture

Culture influences how people perceive situations and organize institutions (Archambault & Archambault, 2003). Cultural differences between nations are an important factor that influences the reporting and disclosure behavior with regard to financial statements (Alexander, Britton, & Jorissen, 2005). According to Zehri and Chouaibi (2013), culture is a major critical factor in explaining the choice of relevant accounting systems that are appropriate for each nation. For example, Zeghal and Mhedhbi (2006) find that developing countries, which belong to an Anglo-American culture, are the most motivated to adopt IFRS.

According to Rezaei (2013), a huge portion of non-U.S.-based capital markets is excluded from the financial community due to a language barrier and current financial analysis software suites being language dependent. In general, business reports are prepared in only one language and mostly in English. English is the most commonly used language on the Internet, and it is estimated that over 50% of all websites are written in English (Wunnava and Leiter, 2008). Wunnava and Leiter (2008) confirm that English is a very significant determinant of Internet adoption. XBRL can overcome the language barrier because it can support 120 languages. Even if a human user does not understand the language used in the financial report, an XBRL-ready computer application can translate and analyze the financial reports automatically for the user (Rezaei, 2013). Hence, the adoption of XBRL can facilitate transactions. For example, changing the language from English to German or Polish can enhance comparability (Shanmuganathan, 2016). Therefore, we can estimate that countries that have an Anglo-American culture are more likely to adopt XBRL. Accordingly, we hypothesize that:

H1: Countries that have an Anglo-American culture are more likely to adopt XBRL.

3.2. Strength of investor protection

According to La Porta et al. (2000), the key mechanism for protecting outside investors, whether they are shareholders or creditors, is the legal system. This mechanism means both laws and their enforcement. When investor rights, such as the shareholders' voting rights and the creditors' reorganization and liquidation rights, are extensive and well enforced by regulators and the courts, investors are willing to finance companies. In contrast, when the legal system does not protect outside investors, corporate governance and external finance do not work well.

Using a sample of 49 countries, La Porta et al. (1997) investigate the link between the legal environment and the capital markets. They find that the legal environment in different countries as measured by both the legal rules and the quality of law enforcement has an important effect on the companies' abilities to obtain more external funding. They find that common law countries have the best protection for investors and benefit from better access to external financing than civil law countries, particularly French civil law countries. Given the benefits of XBRL, countries that protect investors are encouraged to adopt XBRL. On the other hand, countries with weak investor protection laws are also motivated to adopt XBRL in order to improve their investor protection and to gain a better reputation and, then, attract investors.

According to Baldwin, Brown, and Trinkle (2006), a country tends to adopt XBRL to ensure the production of relevant information for decision-making, since it contributes to improving the quality of financial reporting. The country sees XBRL as a method to improve the transparency of financial information that protects investors. This protection stems from the availability of useful, complete, and timely disclosed information since investors must be able to analyze financial information. Consequently, XBRL makes this information easier to understand and easier to process that then facilitates investors' decision-making. Thus, our second hypothesis is:

H2: Investor protection affects the likelihood of XBRL adoption.

3.3.Burden of government regulation

The burden of government regulation refers to how burdensome is it for a country's businesses to comply with governmental administrative requirements (e.g., permits, regulations, and reporting) (WEF, GCR 2011-2012). Government agencies require companies to publish financial information regularly for many different purposes. They require such information to verify that the companies are in the process of meeting the established laws and regulations (Sinnett & Willis, 2009; Bharosa, Janssen, & Winne, 2012). However, these requirements represent an administrative burden on all companies around the world (OECD, 2009; Chen, 2012).

According to Watson (2010, p.13), "the lack of a common reporting standard/language makes these burdens onerous, as business have to tailor, duplicate and individually supply information to meet the many and varied requirements imposed by domestic and international government agencies and regulators.” This burden may even have a significant impact on the country’s economy (Chen, 2012). In order to avoid these problems, several countries have adopted Standard Business Reporting (SBR)/XBRL. SBR is a program that is used to reduce the burden of disclosure. The SBR program is based on the XBRL that is a national financial taxonomy that companies use to report financial information to the government (OECD, 2009). This taxonomy seeks to avoid unnecessary or duplicated data descriptions (OECD, 2009) and the standardization of the business and financial terms in the reporting (Chen, 2012; Richards, Smith, & Saeedi, 2007 ; Chang & Jarvenpaa, 2005). In addition, once created, companies can use the XBRL data several times with no problem that greatly reduces the overall cost of the creation of the data and observation of compliance obligations (Bharosa, Janssen, & Winne, 2012; OECD 2009; Chang & Jarvenpaa, 2005). The XBRL also reduces the costs of preparing and producing reports by eliminating the time and effort required to re-key and verify information (Richards, Smith, & Saeedi, 2007; XBRL France, 2006; Vasal & Srivastava, 2002; Willis, Tesnière, & Jones, 2002). The software programs can perform automatic validation and detect fraud much faster than a manual approach and, thus, reduce labor costs (Chen, 2012).

The Netherlands is the pioneer in SBR since it found that, on average, the government could reduce its demand for some 200,000 pieces of information to only 8000 pieces (Alles, 2009). Therefore, the Netherlands’ companies could save $350 million per year (OECD, 2009). Therefore, governments in different countries may encourage innovation like XBRL. Thus, our third hypothesis is:

H3: A country that has a greater burden of government regulation is more likely to adopt XBRL.

3.4.The level of economic development

Economic conditions are major determinants in the development of a country’s accounting and financial system (Zeghal & Mhedhbi, 2006). Adhikari and Tondkar (1992) consider that a country's stage of development has an influence on its accounting development and practice. As economies develop, the social function of accounting to measure and communicate economic data becomes much more important (Adhikari & Tondkar, 1992). According to Othman and Zeghal (2008), in countries where the level of economic growth is relatively high, the social function of accountancy as an instrument of measurement and communication is of considerable importance. Business and economic activities reach a size and complexity that require sophisticated, high-quality corporate disclosure practices.

Thus, we estimate that the adoption of XBRL increases with the level of economic development. Wunnava and Leiter (2008) indicate that richer countries have well-developed market economies and well-established legal systems. Consequently, they are able and willing to invest more in research and development, and innovation. Thus, our fourth hypothesis is:

H4: A country that has a greater level of economic development, is more likely to adopt XBRL.

3.5.The degree of external economic openness

Foreign investors, international accounting firms, international credit rating agencies, multinational corporations, and world financial institutions represent external pressures (Zeghal & Mhedhbi, 2006; Othman & Zeghal, 2008). They have the potential to influence companies' decision to adopt XBRL. According to Zehri and Chouaibi (2013), while a country's openness to the outside world enhances economic growth, it also engenders greater risks in terms of the security and scale of international pressure. These pressures are reflected in the volume of economic affairs. Such pressures can lead some countries to adopt XBRL. For example, according to Hamid and Salleh (2005), foreign investors are usually more cautious in protecting their economic interests in foreign companies. Through XBRL, foreign investors can effectively monitor their interests.

According to Cooke and Wallace (1990), the more a country's economy is open to the outside world, the more the country is exposed to external pressures. Thus, our fifth hypothesis is:

H5: A country that has a higher degree of external economic openness is more likely to adopt XBRL.

3.6.Macro economic stability

Macroeconomic stability encourages companies and investors to participate in the stock market (Yartey, 2008); this is because of the predictability of the investment environment (Kemboi & Tarus, 2012). By contrast, a highly volatile environment discourages companies and investors from participating in the market. According to Agustini (2016), a country's macroeconomic conditions affect transactions either in the stock market or in the capital market.

Inflation is one factor that contributes to price changes; therefore, there is often a discrepancy between the price of the goods or services obtained the first time and the price afterwards (Agustini, 2016). Conditions of high inflation increase the risk of investment projects and encourage speculative investment, failure of development, and economic instability. Consequently, companies must be able to consider a strategy for when inflation is increasing (Agustini, 2016).

According to Archambault and Archambault (2003), companies that operate in high inflation environments may increase disclosure to further assist investors. Improved comparability with financial reporting in other countries may be of more value in an inflationary environment (Archambault & Archambault, 2009). Therefore, our sixth hypothesis is:

H6: Countries that have high inflation rates are more likely to adopt XBRL.

3.7.Education

The education factor plays an essential role in the use and adoption of innovations (Rawashdeh, Selamat, & Abdullah, 2011). Education includes knowledge, basic skills, and the required confidence to operate innovation-related applications successfully (Hambrick & Mason, 1984; Troshani & Doolin, 2005). High-quality scientific research institutions can generate the basic knowledge needed to build new technologies (WEF, GCR-2011- 2012). According to Pinsker (2005), business graduates need to know how to use XBRL in order to be successful in their jobs. Tertiary education institutions, such as universities, are well-placed to disseminate knowledge about XBRL, its benefits, and how to use it (Troshani & Doolin, 2007). According to Troshani and Doolin (2005), employees should have a basic understanding of XBRL's functionality, benefits, and applications. Pinsker (2003) indicates that knowledgeable graduates are more likely to contribute to the enlargement of the scope of the adoption and implementation of XBRL by their future employers. Warren (2004) finds a positive relation between education and the propensity to adopt technology. According to Rahwani (2013), XBRL education is one of the essential factors for a country to successfully implement XBRL. Based on these discussions, our seventh hypothesis is:

H7: A country that has a more highly educated population is more likely to adopt XBRL.

3.8.Country's firm-level technology absorption

Absorptive capacity refers to a firm's ability to evaluate and utilize outside knowledge (Cohen & Levinthal, 1990). Absorptive capacity not only refers to the acquisition of new knowledge and technologies, but also the potential for exploitation (Cohen & Levinthal, 1990). Managers must be able to learn how to use the technology for it to be successfully adopted (Pinsker, 2008). Generally, companies with high levels of absorptive capacity will have the capacity to acquire and assimilate new technologies quickly because managers of these companies can easily learn technology. Pinsker (2007) affirms that more the technology is easy to be learned, if there is a need for it, more the company will adopt it.

Therefore, companies that have high levels of absorptive capacity, have excellent ability to acquire and exploit new knowledge, and subsequently are more likely to adopt XBRL compared to companies with lower levels of absorptive capacity. A study conducted by Pinsker (2008) in the American context shows a significant and a positive association between the absorption capacity and the adoption of XBRL. Therefore, our eighth hypothesis is:

H8: A country with greater level of technology absorption is more likely to adopt XBRL.

3.9. The accounting system : IFRS

According to Zehri and Chouaibi (2013), more than 100 countries have adopted or expressed their intention to adopt or to converge towards the IFRS.A number of researchers (e.g., Sudalaimuthu & Haraiharan, 2011; Zeghal & Mhedhbi, 2012; Othman & Kossentini, 2015; Turki, Wali, & Boujelbene, 2016) confirm the benefits of the adoption of IFRS. For example, Turki, Wali, and Boujelbene (2016) examine whether the mandatory adoption of IFRS/IAS is beneficial in terms of the information content of earnings. Their results show that the information content of earnings improves after the mandatory adoption of IFRS, and this improvement is reflected in a reduction in capital cost and the errors and dispersion of financial analysts' forecasts. Their sample consists of all listed French companies in the CAC All Tradable Index for the period from 2002 to 2012.

Moreover, Othman and Kossentini (2015) investigate the country-level association between the extent of IFRS adoption and the development of emerging stock markets. Their analysis is based on a sample of 50 emerging economies over a period spanning from 2001 to 2007. They find that a higher level of IFRS adoption has both a positive and significant effect on stock market development.

Therefore, we expect that countries that adopt IFRS to enhance their financial statements will be more willing to adopt XBRL. XBRL and the IFRS is a win-win situation

(Shanmuganathan, 2016). Accoding to Sudalaimuthu & Haraiharan (2011), IFRS and XBRL are two different projects; however, a combined project implementation approach can enable greater efficiency and control over reporting. Both IFRS and XBRL are intended to standardize financial reporting in order to promote transparency and to improve the quality and comparability of business information. Therefore, the two form a perfect partnership. Accordingly, our ninth hypothesis is:

H9: A country that adopts IFRS is more likely to adopt XBRL.

4. Methodology

4.1. Sample selection

Our objective is to explain a country's decision on whether to adopt or not adopt XBRL. In order to achieve this objective, we selected a large number of countries based on the list provided by the Standard & Poor's (2014) website. We excluded from the sample countries with missing data. Our sample consists of 71 countries (please refer to Table 1 and Appendix 1) and covers the period from 2009 to 2014. China was the first country to officially adopt XBRL in 2004 (Wang, Wen, & Seng, 2014). However, we chose to start in 2009 and finish in 2014 due to the high number of missing values for many variables. Indeed, only a few countries had adopted XBRL before 2009 (e.g., in 2004, only two countries had adopted XBRL).

Insert Table 1 here

4.2. Data collection

We collected the data related to our dependent variable from several sources. These include: (i) published articles (e.g., Peng, Shon, & Tan, 2011; Markelevich, Shaw, & Weihs, 2015; Yoon, Zo, & Ciganek, 2011; Rahwani, 2013; Chen, 2012; Dufresne, 2010); (ii) the website of XBRL international consortium (2013) (www.xbrl.org); and (iii) other websites (www.sca.gov.ae; www.xbrl.ca; www.mca.gov.in/xbrl; www.egov-estonia.eu/xbrl). We also used reports published by members of the XBRL international consortium (e.g., O'Kelly, 2010; Sakuta, 2008; Wada, 2006; Pasmooiji, 2010; Maguet, 2012). We also used a report prepared for the Association of Chartered Certified Accountants (ACCA) and the International Association for Accounting Education and Research (IAAER) in support of the work of the International Auditing and Assurance Standards Board (IAASB) by Debreceny, Farewell, and Verkruijsse (2012). Finally, we used a report published by PWC (2013) by Ernst and Young (2010) and other sources (see Appendix 1). For our independent variables, we collected data from World Development Indicators (WDI) on the World Bank website (http://data.worldbank.org) and from the Global Competitiveness Report (from 2009 to 2014) of the World Economic Forum (WEF) and Deloitte's website (www.ias.plus).

4.3 Econometric modeling

Since our dependent variable is dichotomous, we use a logistic regression to analyze the determinants of XBRL adoption. Our empirical model is as follows:

$$\begin{aligned}\mathrm{ADOPXBRL}_{it} &= \alpha_0 + \alpha_1\mathrm{CULT}_{it} + \alpha_2\mathrm{INVPRO}_{it} + \alpha_3\mathrm{BGR}_{it} + \alpha_4\mathrm{LED}_{it} + \alpha_5\mathrm{DEEO}_{it}\\ &\quad + \alpha_6\mathrm{MES}_{it} + \alpha_7\mathrm{EDUC}_{it} + \alpha_8\mathrm{TAB}_{it} + \alpha_9\mathrm{ACCSYS}_{it} + \mathrm{YEAR\ EFFECTS} + \varepsilon\end{aligned}$$

Where:

ADOPXBRL: The adoption of XBRL

We used the following four ways to measure our dependant variable, namely, the adoption of XBRL:

1. One if country (i) adopts XBRL without considering the purpose of adoption and zero otherwise.

2. One if country (i) adopts XBRL to meet the information needs of securities regulators and zero otherwise.
3. One if country (i) adopts XBRL to meet the information needs of at least one other government agency (chamber of commerce, companies house, statistical offices, tax administration, central balance sheet, banking regulators) and zero otherwise.
4. One if country (i) adopts XBRL to meet the information needs of both securities regulators and at least one other government agency and zero otherwise.

The following are the definitions of our independent variables:

CULT: Culture is a dummy variable that takes the value one if the country has an Anglo-Saxon culture and zero otherwise (source: specific country-based information).

INVPRO: Strength of investor protection is a combination of the extent of the disclosure index (transparency of transactions), the extent of director liability index (liability for self-dealing), and the ease of the shareholder suit index (shareholders' ability to sue officers and directors for misconduct). It is measured on a 0–10 (best) scale (source: The Global Competitiveness Report of the World Economic Forum, WEF).

BGR: Burden of government regulation is measured by a score of one [extremely burdensome] to seven [not burdensome at all] (source: The Global Competitiveness Report of the World Economic Forum, WEF). We have reversed this order of the burden of government regulation score in our regressions to make it easier to interprete. Thus, one is now the lowest burden of government regulations and seven is the highest burden of government regulations.

LED: The level of economic development is measured by the annual growth rate of the gross domestic product (GDP) per capita (source: World Development Indicators, WDI, World Bank).

DEEO: Degree of external economic openness is measured by the average of six years of the net inflows of foreign direct investment, divided by the GDP (source: World Development Indicators, WDI, World Bank).

MES: Macroeconomic stability is measured by the average inflation rate (source: World Development Indicators, WDI, World Bank).

EDUC: Education is measured by the gross tertiary education enrollment rate (source: The Global Competitiveness Report of the World Economic Forum, WEF).

TAB: Country's firm-level technology absorption is measured by a score of 1 to 7 [1 = not at all; 7 = adopt extensively] (source: The Global Competitiveness Report of the World Economic Forum, WEF).

ACCSYS: Accounting system is represented by the adoption of IFRS. The adoption of IFRS is measured by a dummy variable that takes the value of one if country (i) requires the adoption of IFRS and zero otherwise (source: Deloitte's website, www.ias.plus).

Year effects represent year dummy variables.

And ε is the margin of error.

Index i indicates the individuals (the countries), and t is for time (the period covers the years from 2009 to 2014).

5. Empirical findings

Table 2 presents the mean, standard deviation, the minimum, and the maximum for the independent variables. The sampled countries show a very large variability in all these variables. The variable strength of investor protection ranges from 2.7 to 9.7 with a mean of 5.801 and a standard deviation of 1.481. Burden of government regulation ranges from 3 to 7.9 with a mean of 5.607 and a standard deviation of 0.936. The degree of external economic openness varies widely among the 71 countries with a minimum of -15.989 and a maximum of 252.308. There is a very large variation between the countries in terms of the level of economic development, macroeconomic stability, education and the country's firm-level technology absorption. For the variable culture and accounting system, there are 45.07% country-year observations with an Anglo-Saxon culture and 78.87% country-year observations in which IFRS is mandatory.

Insert Table 2 here

With regard to the adoption of the variable XBRL, Table 3 shows that 69.25% of the country-year observations show no adoption while 30.75% of the country-year observations show adoption.

XBRL is adopted to meet the information needs of several government agencies such as securities regulators; this ensures the quality of information provided to information users, investors, and analysts. The adoption of XBRL also means that the information needs of chambers of commerce, companies house, statistical offices, tax administration, central balance sheet, and banking regulators can be met. For 16.90% of the country-year observations, the XBRL is adopted to meet the information needs of securities regulators. Of the country-year observations, 21.83% show that XBRL is adopted to meet the information needs of other government agencies. Further, 7.98% of the country-year observations show that XBRL is adopted to meet both the securities regulators' and other government agencies' needs.

In general, the number of countries that adopt XBRL vary between 16 in 2009 to 26 in 2014. However, while the adoption is increasing, just a few countries have adopted XBRL so far. This may be due to the lack of knowledge about XBRL and its benefits.

Insert Table 3 here

We examine the difference between the countries that adopted XBRL and those that did not by using a parametric Student test for the quantitative variable (the level of economic development) and the non-parametric Mann-Whitney test for the other independent variables. The choice of one test rather than another is based on the normality test (the Kolmogorov-Smirnov test). The level of economic development validates the hypothesis of normality. We also used the Mann-Whitney test for the dummy explanatory variables (culture and accounting system).

The results show that the level of economic development does not significantly differ between the countries that adopted XBRL and those that did not (see Table 4: Panel A).

Table 4 (Panel B) presents the results of the Mann-Whitney test. These show significant differences in the strength of investor protection, importance of equity market, macroeconomic stability, education and the country's firm-level technology absorption at the 1% level between the countries that adopted XBRL and those that did not. For culture, burden of government regulation, degree of external economic openness, and accounting system, there is no significant difference between the two groups.

Insert Table 4 here

Table 5 shows the correlation analysis. It shows that the adoption of XBRL is significant at the 1% level and positively associated with the strength of investor protection, the degree of external economic openness, the education-level and the country's firm-level technology absorption. Also, the burden of government regulation is negative and significant at the 5% level. Therefore, a higher burden of government regulation decreases the probability of adopting XBRL. Macroeconomic stability is negative and significant at the 1% level. Hence, countries that have higher levels of inflation are less likely to adopt XBRL. The table also shows that multicollinearity is not a problem in our model. In addition, we conducted the Variance Inflation Factor (hereafter refered to as VIF) test. All the VIF values of our independent variables are less than the critical value of 5. Thus, multicollinearity is not a concern in our analysis.

Insert Table 5 here

Table 6 shows the determinants of XBRL adoption. It shows a number of findings. First, we find that culture is negatively associated with the adoption of XBRL. Hence countries that have an Anglo-Saxon culture are less likely to adopt XBRL. These countries tend to overcome the language barriers and thus attract more investors. This is because XBRL improves and speeds up access to information in all languages and improves transparency. Further, foreign investors can rely on information in XBRL-tagged financial reports to make investment decisions without having to translate the financial statements into their language (Markelevich, Shaw, & Weihs, 2015). Therefore, our first hypothesis, H1, is rejected.

Second, investor protection has a positive and significant effect (at the 1% level) on the adoption of XBRL. Thus, the probability of a country adopting XBRL increases with the level of investor protection. Therefore, our second hypothesis, H2, is accepted. According to Chung (2006), a good investor protection environment minimizes the costs of information asymmetry and reduces the probability of trading against informed traders. In contrast, a weaker investor protection environment leads to greater expropriation by managers and higher asymmetric information costs. XBRL helps to improve the transparency; to reduce the asymmetry of information and the earnings management; and to improve the firm's governance (Birt, Muthusamy, & Bir, 2017; Shan & Troshani, 2016; Shkurti & Allko, 2016; Liu, Wang, Yao, 2014; Bai, Sakaue, & Takeda, 2014; Wang & Gao, 2012; Peng, Shon, & Tan, 2011; Ernst & Young, 2010; SEC, 2009; Tan & Shon, 2009 ; Premuroso & Bhattacharya, 2008; Li et al., 2006; Hannon, 2005; Shin, 2003; Willis, Tesnière, & Jones, 2002; Debreceny & Gray, 2001).

Similarly, the level of economic development positively and significantly (at the 1% level) affects the countries' decision to adopt XBRL. Literature shows that the level of economic development constitutes a major determinant in adopting IFRS (e.g., Zehri & Chouaibi, 2013; Zeghal & Mhedhbi, 2006). IFRS and XBRL tend to enhance the quality of information (e.g. Shanmuganathan, 2016; Shan &Troshani, 2016; Markelevich, Shaw, & Weihs, 2015). Hence, we can conclude that countries that have well-developed economies are more likely to adopt XBRL. We, therefore, accept H4. Regarding the degree of external economic openness, it is positive and significant (at the 1% level). Consequently, we conclude that the more the country is open to the outside word, the more likely it will adopt XBRL. Hence, H5 is accepted. We find also that the variable education has a positive and significant (at the level of 1%) impact on the decision to adopt XBRL. This indicates that the higher the education level of a country, the more likely the country will adopt XBRL. Therefore, H7 is supported. In addition, the country's firm-level technology absorption is found to be significantly (at the level of 1%) associated with the adoption of XBRL. This suggests that countries with high levels of technology absorption are more likely to adopt XBRL. Therefore, H8 is accepted.

However, the burden of government regulation, the macroeconomic stability and the mandatory adoption of IFRS have no significant influence on the decision to adopt XBRL. We therefore reject H3, H6 and H9. The year dummies are insignificant for all years.

Insert Table 6 here

5.1. Comparaison between the different purpose of adoption of XBRL: The determinant of adoption of XBRL to meet the information needs of securities regulators, of other government agencies and of both securities regulators and other government agencies

We measure our dependant variable, the adoption of XBRL, by using the following three conditions:

1. One if country (i) adopts XBRL to meet the information needs of securities regulators and zero otherwise.

2. One if country (i) adopts XBRL to meet the information needs of at least one of the other government agencies and zero otherwise.

3. One if country (i) adopts XBRL to meet the information needs of both securities regulators and at least one of other government agencies and zero otherwise.

Table 7 presents the pooled logistic regression results. For the adoption of XBRL to meet the information needs of securities regulators, we find that the strength of investor protection, the level of economic development, the education-level and the country's firm-level technology absorption have positive and significant effects (at the level of 1%) on the decision to adopt the XBRL. Thus, countries that have these characteristics are more willing to adopt XBRL to meet the information needs of securities regulators. However, the variable culture has a negative and significant coefficient (at the 5% level). The variables burden of government regulation, degree of external economic openness and macroeconomic stability are insignificant. The year dummies are insignificant for all the years except 2010.

Next, for the adoption of XBRL to meet the information needs of at least one of the other government agencies, we find that the burden of government regulation is statistically significant at the 1% level and has a positive effect on the adoption of XBRL. Therefore, we can conclude that the greater the administrative burdens, the greater the motivation for the country to adopt XBRL to meet the information needs of government agencies. Companies can use XBRL to save costs and streamline their processes for collecting and reporting financial information (Bharosa, Janssen, & Winne, 2012; Financial Reporting Council, 2009). The degree of external economic openness is positive and significant (at the 1% level). Therefore, external pressures can affect the decision to adopt XBRL to meet the information needs of government agencies. Similarly, the variables of level of education and level of technology absorption are positive and have a significant (at the 1% level) influence on the decision to adopt XBRL. This result confirms that countries with a higher education levels and higher levels of technology absorption are more likely to adopt XBRL. However, the variables of culture, strength of investor protection, level of economic development, macroeconomic stability and the mandatory adoption of IFRS have insignificant impact on a country's decision to adopt XBRL to meet the information needs of at least one other government agency. The year dummies are insignificant for all years except 2012.

To meet the information needs of both securities regulators and at least one other government agency, the burden of government regulation, the education-level and the level of technology absorption have positive and significant impacts (at the 1% level) on a country's decision to adopt XBRL. These findings show that countries with a greater burden of government regulation, higher education levels and higher levels of technology absorption are more inclined to adopt XBRL. Contrary to our predictions, the variable mandatory adoption of IFRS is significant (at the 5% level) and negative. Countries should use other policies that are complementary to the adoption of IFRS to promote the adoption of XBRL to meet the information needs of both securities regulators and other government agencies. The variables of culture, investor protection, level of economic development, degree of external economic openness and macroeconomic stability do not affect a country' decision to adopt or not XBRL. The year dummies are insignificant for all years.

Insert Table 7 here

5.2. Comparison of the determinants of adoption of XBRL between developed and developing countries

We examine the factors that explain the developed countries' and developing countries' individual decisions to adopt XBRL. We use a pooled logistic regression. Our dependant variable is one if country (i) adopts XBRL without considering the purpose of adoption and zero otherwise. Our explanatory variables are the culture, the strength of investor protection, the burden of government regulation, the level of economic development, the degree of external economic openness, macroeconomic stability, education, the country's firm-level technology absorption and the mandatory adoption of IFRS.

We split our sample into two subsamples: one for developed countries and one for developing countries (see Table 8). For the developed countries, we find that the burden of government regulation positively and significantly (at the 10% level) affects the country's decision to adopt XBRL. As for the level of economic development, it is positive and significant (at the 10% level). In addition, the degree of external economic openness has a positive and significant (at the 10% level) impact on the adoption of XBRL. The significance of these variables indicate that developed countries with a heavy burden of government regulation, higher levels of economic development and higher degree of external economic openness are more likely to adopt XBRL. The variables for culture, strength of investor protection, macroeconomic stability, education, country's firm-level technology absorption and IFRS have insignificant impacts on the adoption of XBRL. Also, the year dummies are insignificant for all years.

For developing countries, the variables for culture, strength of investor protection, level of economic development, degree of external economic openness, macroeconomic stability, country's firm-level technology absorption and IFRS have a significant impact on the decision to adopt XBRL. The cofficient for culture is negative and significant at the 1% level. Hence, developing countries with an Anglo-American culture are less likely to adopt XBRL. The variables for investor protection, level of economic development, country's firm-level technology absorption and the mandatory adoption of IFRS positively and significantly (at the level of 1%) affect the adoption of XBRL. Thus, the probability of a developing country adopting XBRL increases with the strength of investor protection, the greater levels of economic development, the high levels of technology absorption and with the mandatory adoption of IFRS. In addition, the variable macroeconomic stability has a positive and significant effect (at the 5% level) on the adoption of XBRL. Hence, the higher the level of inflation, the more likely a developing country will adopt XBRL. We find that countries with a less stable macroeconomic environment are more likely to adopt XBRL to encourage investors to participate in their stock market. The degree of external economic openness has both a negative and significant effect at the 5% level on whether or not to adopt XBRL. This is inconsistent with our prediction that countries with a high degree of external economic openness are more likely to adopt XBRL. This finding indicates that the degree of external economic openness does not affect developing countries's decision to adopt XBRL. Thus, these countries should implement other policies that are complementary to external economic openness to promote the adoption of XBRL. However, the variables for the burden of government regulation and education do not have a significant impacts on the decision to adopt XBRL. In addition, the year dummies are insignificant for all years except 2011 and 2011.

Insert Table 8 here

As a summary of our analysis, by applying pooled logistic regressions, we conclude that the burden of government regulation, the level of economic development and the degree of external economic openness represent the major determinants as to whether or not developed countries adopt XBRL. Furthermore, the strength of investor protection, the level of economic development, the macroeconomic stability, the country's firm-level technology absorption, and the mandatory use of IFRS have positive impacts on the adoption of XBRL in developing countries. However, the variables of culture and degree of external economic openness negatively affect the developing countries' decisions to adopt XBRL.

Additional analysis

In our main analysis, we used pooled logistic regressions. In this section, we examine each year separately. We use binary logistic regressions.

Table 9 presents the results of the determinants of the adoption of XBRL to meet the information needs of securities regulators or other government agencies. The analysis for year 2009 shows that the strength of investor protection and the level of technology absorption have a positive and significant (at the level of 5% and 1%) effect on the countries' adoption of XBRL. The culture variable is both negative and significant (at the 5% level). Hence, countries with strong investor protection and high levels of technology absorption are more likely to adopt XBRL. In contrast, the countries with an Anglo-Saxon culture are less inclined to adopt XBRL. For year 2010, the variables of education and technology absorption have a positive and significant (at the level of 5%) effect on a country' decision to adopt XBRL. For year 2011, the level of technology absorption impacts positively and significantly (at the 1% level) the country's decision to adopt XBRL. For the years 2012 and 2013, the strength of investor protection and the level of technology absorption are both positive and significant. Finally, the strength of investor protection, the level of economic development and the level of technology absorption have a positive and significant effect on the adoption of XBRL in year 2014.

Insert Table 9 here

For the adoption of XBRL to meet the information needs of securities regulators, Table 10 shows that for the year 2009, the strength of investor protection and the country's firm-level technology absorption have both a positive and significant effect (at the 5% level) on the adoption of XBRL. For 2010, there is no significant result. For the year 2011, both the strength of investor protection and the level of technology absorption are positively and significantly (at the level of 10% and 5%) related to the decision to use XBRL. For 2012, the country's firm-level technology absorption has a positive and significant (at the 5% level) effect on the adoption of XBRL. For 2013, the variables of burden of government regulation, level of economic development and level of technology absorption affect positively and significantly the decision to adopt or not XBRL. For year 2014, the strength of investor protection, the macro economic stability and the country's firm-level technology absorption are positively and significantly (at the 10% level) related to the adoption of XBRL.

Insert Table 10 here

For the adoption of XBRL to meet the information needs of at least one other government agency, Table 11 shows that for the year 2009, there is no significant result. For 2010, the degree of external economic openness, the education and the country's firm-level technology absorption have a positive and significant effect on the decision to adopt XBRL. Therefore, countries that have high degree of external economic openness, higher education-levels and higher levels of technology absorption are more likely to adopt XBRL. For 2011, the education-level and the country's firm-level technology absorption affect positively and significantly (at the level of 5%) the decision to adopt XBRL. For 2012, the variable education is positive and significant (at the level of 5%). For 2013, we find that the the probability of a country adopting XBRL increases with the level of the burden of government regulation and the level of education. For 2014, the level of economic development and the education variables are positive and significant (at the level of 5%). Hence, countries with higher levels of economic development and a higher levels of education are more likely to adopt XBRL.

Insert table 11 here

6. Conclusion

The objective of this study is to identify the key factors associated with the countries' decisions to adopt XBRL. We use the pooled logistic regressions on 71 countries over the period from 2009 to 2014.

The results of the empirical analysis indicate that the countries that are most likely to adopt XBRL are those that have strong investor protection, higher levels of economic development, higher degree of external economic opennes, more highly educated populations, and high levels of technology absorption. We find this result by measuring the dependant variable, adoption of XBRL, by one if country (i) adopts XBRL without considering the objective of that adoption and zero otherwise.

Then, we use three other measures for the adoption of XBRL. This first measure is one if country (i) adopts XBRL to meet the information needs of securities regulators and zero otherwise. The second measure is one if country (i) adopts XBRL to meet the information needs of at least one other government agency (which are the chamber of commerce, companies house, statistical offices, tax administration, central balance sheet, and banking regulators) and zero otherwise. The third measure is one if country (i) adopts XBRL to meet the information needs of both securities regulators and at least one other government agencies. Using each measure, we find a number of factors affecting the country's decision to adopt XBRL. In addition, we investigate the separate factors that affect developing and developed countries' decisions to adopt XBRL without considering the purpose of adoption. Our findings show that the burden of government regulation, the level of economic development and the degree of external economic opennes positively and significantly affect developed countries' decisions to adopt XBRL. For developing countries, culture, investor protection, level of economic development, degree of external economic openness, macroeconomic stability, country's firm-level technology absorption and the mandatory adoption of IFRS affect the adoption of XBRL. Our findings show that developing countries, which have an Anglo-Saxon culture, are less likely to adopt XBRL. Moreover, developing countries with strong investor protection, higher levels of economic development, high inflation rates, high levels of technology absorption, and mandate the adoption of IFRS are more likely to adopt XBRL. In contrast, the degree of external openness negatively affects the decision to adopt XBRL.

We test all these models with a pooled logistic regression. In addition, we test these models for each year. We find a number of factors that contribute to a country's decision to adopt XBRL.

This study has some limitations. We find some problems regarding the collection of some information: countries that adopt XBRL and those that do not adopt; the purpose and the year of adoption. While in this research we tried to be accurate in using all this information these difficulties could be eliminated in future research and all this information could be more available and accurate. Also, only a small number of countries have adopted XBRL. Therefore, it would be interesting to repeat the work when more countries have adopted XBRL.

This paper provides some policy implications. In order to attract foreign investment, it is important to improve information quality through the adoption of XBRL. Hence, it is essential that regulators facilitate the adoption of XBRL. Our study is especially useful for researchers, professionals, and countries that have not yet adopted XBRL to better understand the specific factors that may promote the adoption of XBRL. It also helps the XBRL international consortium and regulators to formulate effective strategies that facilitate all countries to adopt XBRL.

This paper suggests a number of avenues for future research. It would be interesting to consider other determinants that may affect the countries' adoption of XBRL. Also, it would be interesting to investigate the consequences of the adoption of XBRL at the country level.

References

Agustini, A. T. (2016). The effect of firm size and rate of inflation on cost of capital: The role of IFRS adoption in the world. *Social and Behavioral Sciences*, *219*, 47-54.

Alles, M. (2009). Special issue: XBRL and the future of disclosure. *International Journal of Disclosure and Governance*, *6* (3), 184-187.

Arnold, V., Bedard, J.C., Phillips, J. R.., & Sutton. S. G. (2012). The impact of tagging qualitative financial information on investor decision making: Implications for XBRL. *International Journal of Accounting Information Systems, 13*(1), 2-20.

Adhikari, A., & Tondkar, R. H. (1992). Environmental factors influencing accounting disclosure requirements of global stock exchanges. *Journal of International Financial Management & Accounting*, *4*(2), 75–105.

Alexander, D., Britton, A., & Jorissen, A. (2005). International Financial Reporting and Analysis. 2nd edition. London: Thomson.

Archambault, J. J., & Archambault, M. E. (2003). A multinational test of determinants of corporate disclosure. *International Journal of Accounting*, *38*(2), 173–194.

Archambault, J. J., & Archambault, M. E. (2009). An analysis of social factors influencing the adoption of international financial reporting standards of international financial reporting standards. *Journal of Global Business Advancement*, *2* (1-2), 38-53.

Bahmani, D. (2014). The relation between disclosure quality and information asymmetry : Empirical evidence from Iran. *International Journal of Financial Research*, *5*(2), 110–114.

Bai, Z., Sakaue, M., & Takeda, F. (2014). The impact of XBRL adoption on the information environment: Evidence from Japan. *The Japanese Accounting Review, 4,* 49-74.

Baldwin, A., Brown, C., & Trinkle, B. (2006). XBRL: An impacts framework and research challenge. *Journal Of Emerging Technologies In Accounting*, *3*, 97-116.

Banque Nationale de Belgique. (2006). La taxonomie XBRL des comptes annuels normalisés.

Notice explicative de la Centrale des bilans. Notice explicative de la Centrale des bilans. Retrieved from https://www.nbb.be/doc/ba/xbrl/taxo2017/userguide/xbrl_userguide_taxonomy_v11_f.pdf.

Beilock, R., & Dimitrova, D. V. (2003). An exploratory model of inter-country Internet diffusion. *Telecommunications Policy*, *27*(3–4), 237–252.

Bharosa, N., Janssen, M., & Winne, N. De. (2012). Principles for transforming to Standard Business Reporting : Lessons learned from the Netherlands. *Working Paper*.

Birt, J, L., Muthusamy, K., & Bir, P. (2017). XBRL and the qualitative characteristics of useful financial information. *Accounting Research Journal*, *30* (1), 107-126.

Bonsón. (2001). The role of XBRL in Europe. *The International Journal of Digital Accounting Research*, *1*(2), 101–110.

Bonsón, E., Cortijo, V., & Escobar, T. (2009). A Delphi investigation to explain the voluntary adoption of XBRL. *The International Journal of Digital Accounting Research 9* (15): 193-205.

Boolaky, P., Krishnamurti, C., & Hogue, A. (2013). Determinants of the strength of auditing and reporting standards: a cross-country Study. *Australasian Accounting, Business and Finance Journal*, *7*(4), 17–36.

Boritz, J. E., & Timoshenko L. M. (2015). Firm-specific characteristics of the participants in the SEC's XBRL voluntary filing program, *Journal of Information Systems*, *29*, 9–36.

Chang, C., & Jarvenpaa, S. (2005). Pace of information systems standards development and implementation: The case of XBRL, *Electronic Markets*, *15*(4), 365-377.

Chen, Y. C. (2012). A comparative study of e-government XBRL implementations: The potential of improving information transparency and efficiency. *Government Information Quarterly*, *29*(4), 553–563.

Chen, S., Harris, L., Li, W., & Wu, D. (2015). How does XBRL affect the cost of equity capital? Evidence from an emerging market. *Journal of International Accounting Research*, *14*(2), 123-145.

Chotaliya, M. (2014). Application of XBRL in financial reporting. *Scholars World*, *2*(2), 124–130.

Chung, H. (2006). Investor protection and the liquidity of cross-listed securities: Evidence from the ADR market, Journal of Banking & Finance, *30*, 1485–1505.

Clements, C. E., Neill, J. D., & Scott Stovall, O. (2010). Cultural diversity, country size, and the IFRS adoption decision. *Journal of Applied Business Research*, *26*(2), 115–126.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity : A new perspective on learning and innovation. *Administrative Science Quarterly- Special Issue : Technology , Organizations , and Innovation*, *35*(1), 128–152.

Cooke, T. E., & Wallace, R. S. O. (1990). Financial disclosure regulation and its environment: A review and further analysis. *Journal of Accounting and Public Policy*, *9*, 79–110.

De Brandt, D. (2012). XBRL Belgium: a brief history of time. Retrieved from http://www.eurofiling.info/EIOPAXBRLEuropeSeminarTallin2012/presentations/03.%20XBRL%20Belgium%20a%20brief%20history%20of%20time%20%28D.DeBrandt%29.pdf .

Debreceny, R. S., Chandra, A., Cheh, J. J., Guithues-Amrhein, D., Hannon, N. J., Hutchison, P. D., Janvrin, D., Jones, R. A., Lamberton, B., Lymer, A., Mascha, M., Nehmer, R., Roohani, S., Srivastava, R. P., Trabelsi, S., Tribunella, T., Trites, G., Vasarhelyi, M. A. (2005). Financial reporting in XBRL on the SEC's EDGAR system: A critique and evaluation. *Journal of Information Systems*, *19*,191-210.

Debreceny, R., Farewell, S., Piechocki, M., Felden, C., & Gräning, A. (2010). Does it add up? Early evidence on the data quality of XBRL filings to the SEC. *Journal of Accounting and*

Public Policy, *29*(3), 296–306.

Debreceny, R., Farewell, S., & Verkruijsse, H. (2012). The implications of XBRL for the financial statement audit. A report prepared for the Association of Chartered Certified Accountants (ACCA) and the International Association for Accounting Education and Research (IAAER) in support of the work of the International Auditing and Assurance Standards Board (IAASB). Retrieved from http://files.iaaer.org/research/XBRLAssuranceFinalReport.pdf?1406556395.

Debreceny, R., & Gray, G. (2001). The production and use of semantically rich accounting reports on the internet: XML and XBRL. *International Journal of Accounting Information Systems*, *2*(1), 47-54.

Deloitte. (2011). XBRL The new world of reporting. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/in-risk-xbrl-the-new-world-of-reporting-noexp.pdf.

Dufresne, L. (2010). XBRL and the Belgian ―"Building Blocks" Approach. *XBRL global, 1* (2), 28-31.

Efendi, J., Park, J. D., & Smith, L. M. (2014). Do XBRL filings enhance informational efficiency? Early evidence from post-earnings announcement drift. *Journal of Business Research*, *67*(6), 1099–1105.

Elissavet, S., Antonios, S., & Theodora, L. (2013). Acceptance and usage of extensible business reporting language: An empirical review. *Journal of Social Sciences*, *9*(1), 14-21.

Enachi, M. (2013). XBRL and financial reporting transparency. *Broad Research in Accounting, Negotiation, and Distribution* , *4*(1), 10-19.

Ernst and Young. (2010). Addressing XBRL: Moving business and financial reporting into the digital age. Retrieved from: http://www.ey.com/Publication/vwLUAssets/Addressing-XBRL-Eng/$FILE/Addressing-XBRL-Eng.pdf.

Felo, A. J., Kim, J. W., & Lim, J. H. (2018). Can XBRL detailed tagging of footnotes improve financial analysts' information environment? *International Journal of Accounting Information Systems*, *28*, 45-58.

FFIEC. (2006). Improved Business Process Through XBRL: A Use Case for Business Reporting. Retrieved from http://www.xbrl.org/us/us/FFIEC%20White%20Paper%2002Feb2006.pdf.

Financial Reporting Council. (2009). An Introduction to XBRL. *Working Paper*.

Francis, P. (2012). Intro to XBRL. accountants to day. Retrieved from http://www.mia.org.my/at/at/2012/10/06.pdf .

Ghani, E. K., Said, J., & Muhammad, K. (2014). Enhancing corporate governance via XBRL: preparers' perception on compatibility expectation. *Procedia - Social and Behavioral Sciences, 145,* 308-315.

Gostimir, D. (2015). XBRL standard for financial reporting in Croatia: Current state and perspectives. *Business Systems Research Journal*, *6*(2), 31–40.

Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of Its top managers. *The Academy of Management Review, 9*(2), 193-206.

Hannon, N. (2005). XBRL Fundamentals. *Strategic Finance*, *86*(10), 57–58.

Hope, O.K., Jin, J., & Kang, T. (2006). Empirical evidence on the jurisdictions that adopt IFRS. *Journal of International Accounting Research* , *5*(2), 1-20.Hao, L., Zhang, J. H., & Fang, J. (2014). Does voluntary adoption of XBRL reduce cost of equity capital? *International Journal of Accounting and Information Management , 22* (2), 86-102.

Houqe, M. N., Easton, S., & van Zijl, T. (2014). Does mandatory IFRS adoption improve information quality in low investor protection countries? *Journal of International Accounting, Auditing and Taxation*, *23*(2), 87–97.

IASC Foundation in XBRL Team. (2010). XBRL Update. Retrieved from http://www.ifrs.org/XBRL/Resources/Documents/XBRLUpdateMarch2010.pdf .

Jeanjean, T., Stolowy, H., & Erkens, M. (2012). Economic consequences of adopting English for annual reports. *Working Paper*.

Kaya, D., & Pronobis, P. (2016). The benefits of structured data across the information supply chain: Initial evidence on XBRL adoption and loan contracting of private firms. *Journal of Accounting and Public Policy*, *35*(4), 417–436.

Kemboi, J. K., & Tarus, D. K. (2012). Macroeconomic determinants of stock market development in emerging markets: Evidence from Kenya. *Research Journal of Finance and Accounting*, *3*, 57-68.

Kloeden, P. (2006). XBRL Cost-benefit analysis : In Theory and application. Rochester Institute of Technology RIT Scholar Works. Rochester Institute of Technology. Retrieved from http://scholarworks.rit.edu/article/1627.

Kossentini, A., & Othman, H. (2014). A study of the institutional and economic determinants of IFRS adoption in emerging economies A study of the institutional and economic determinants of IFRS adoption in emerging economies. *Working Paper.*

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (2002). Investor protection and corporate governance. *Journal of Financial Economics*, *58*(1–2), 3–27.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. *The Journal Of Finance, 52*(3), 1131-1150.

Li, O. Z., Lin, Y., & Ni, C. (2011). Does XBRL adoption reduce the cost of equity capital? *Working Paper*.

Li, Y., Roge, J. N., Rydl, L., & Crews, M. (2006). Information technology addresses transparency: the potential effects of XBRL on financial disclosure. *Issues in Information Systems*, *7*(2), 241–245.

Liu, C., Wang, T., & Yao, L. J. 2014. XBRL's impact on analyst forecast behavior: An empirical study. *Journal of Accounting and Public Policy*, *33*(1), 69-82.

Liu, C., Luo, X., & Wang, F. L. (2017). An empirical investigation on the impact of XBRL adoption on information asymmetry: Evidence from Europe. *Decision Support Systems*, *93*, 42–50.

Maguet, G. (2012). The progress of XBRL in Europe. Retrieved from http://www.xbrleurope.org/files/Presentations/20121213-10thXBRLEuropeDayFrankfurt/4-20121213-10thXEUDay-UpdateinEuropeGillesMaguet.pdf .

Maguet, G. (2006). XBRL France, Rapport moral / d'avancement sur l'activité de XBRL France (www.xbrl.fr).

Manmohan, M., & Pk, M. (2014). A study on cost-benefit analysis of XBRL reporting on financial data ecosystem. *International Journal of Advances in Management, Economics and Entrepreneurship*, *1*(1), 1–5.

Markelevich, A., Shaw, L., & Weihs, H. (2015). The Israeli XBRL adoption experience. *Accounting Perspectives*, *14*(2), 117–133.

Nobre, C., & Carvalho, N. (2011). XBRL in Brazil : Current status and prospects, *1*(0), 17–19. Retrieved from http://www.mzweb.com.br/attitudeglobal/web/arquivos/FOA_Brazil_11Aug03.pdf.

OECD. (2009). Forum On Tax Administration : Taxpayer Services Sub-Group. Guidance note Standard Business Reporting. Retrieved from http://www.oecd.org/ctp/administration/43384923.pdf.

O'Kelly, C. (2010). XBRL Adoption Update. Retrieved from http://fr.slideshare.net/xbrlplanet/xbrl-world-wide-adoption-survey-april-2010 .

Othman, H., & Zeghal, D. (2008). A study of corporate governance disclosure and its country-level determinants in the emerging markets. *Research in Accounting in Emerging*

Economies, *8*, 125-155.

Othman, H., & Kossentini, A. (2015). IFRS adoption strategies and theories of economic development: Effects on the development of emerging stock markets. *Journal of Accounting in Emerging Economies*, *5*(1), 70-121.

Pasmooiji, J. (2010). XBRL around the World. eXtensible Business Reporting Language (XBRL)-an Update. Retrieved from http://www.ifac.org/sites/default/files/downloads/Pasmooij_XBRL_Around_the_World.pdf .

Peng, E. Y., Shon, J., & Christine, T. (2011). XBRL and accruals: empirical evidence from China. *Accounting Perspectives , 10*(2), 109–138.

Pinsker, B. R., & Li, S. (2008). Costs and benefits of XBRL adoption: Early evidence. *Communications of the ACM*, *51*(3), 47–50.

Pinsker, R. (2007). A theoretical framework for examining the corporate adoption decision involving XBRL as a continuous disclosure reporting technology. In: Debreceny, R ; Felden, C; Piechocki, M ,ed. New Dimensions of Business Reporting and XBR,. Springer Group: Verlag, Germany.

Pinsker, R. (2005). Teaching XBRL to graduate business students: A hands-on approach. *Journal of STEM Education*, *5*(1), 5–18.

Pinsker, R. (2003). XBRL awareness in auditing: a sleeping giant? *Managerial Auditing Journal , 18*(9), 732-736.

Pinsker, R. (2008). An empirical examination of competing theories to explain continuous disclosure technology adoption intentions using XBRL as the example technology. *The International Journal of Digital Accounting Research , 8*(14), 81-96.

Premuroso, R. F., & Bhattacharya, S. (2008). Do early and voluntary filers of financial information in XBRL format signal superior corporate governance and operating performance ? *International Journal of Accounting Information Systems, 9* (1), 1-20.

PriceWaterHouseCoopers, PWC. (2003). Trusted and efficient financial reporting. *A working paper discussing a new solution to facilitate traight through reporting of business information*. Retrieved from https://www.pwc.com/gx/en/xbrl/assets/pwc_trusted_efficient_financial_reporting.pdf.

PWC. (2013). Global Developments in XBRL. Retrieved from http://macpamedia.org/media/downloads/2013Summit/willis_Global_Developments_XBRL_ppt%202%20per%20page.pdf .

Ragothaman, S. (2012). An empirical analysis of the relationship between corporate governance rating and firms' characteristics: Evidence from MAI Thailand. *International Journal of Economics and Financial Issues*, *3*(1), 50–53.

Rahwani, N. (2013). How to integrate XBRL into Accounting curriculum at Banjarmasin state Polytechnic. *Information Systems International Conference , 2* (4), 1-6.

Rawashdeh, A., & Selamat, M. H. (2013). Critical success factors relating to the adoption of XBRL in Saudi Arabia. *Journal of International Technology and Information Management, 22* (2), 49-70.

Rawashdeh, A. A., Selamat, M. H., & Abdullah, M. S. (2011). Characteristics of consumers influencing adoption behavior for XBRL, *World Review of Business Research*, *1*(3), 1–12.

Rezaei, Z. (2013). XBRL: A new language for financial reporting. *International Research Journal of Applied and Basic Sciences, 6*(4), 437-442.

Richards, J., Smith, B., & Saeed, A. (2007). An introduction to XBRL. *Working Paper*.

Roohani, S., Furusho, Y., & Koizumi, M. (2009). XBRL: Improving transparency and monitoring functions of corporate governance. *International Journal of Disclosure and Governance*, *6*(4), 355–369.

Sakuta, A. (2008). Cases of Real Use in Japan, Financial Services Working Group, XBRL.

Retrieved from http://archive.xbrl.org/18th/sites/18thconference.xbrl.org/files/Regulator-XBRLinAction-Complete.pdf .

Samuel, F. A., Samuel, F. O., & Obiamaka, N. (2013). The impact of international financial reporting standards on taxation. *American Journal of Business Education*, *4*(10), 169–174.

SEC. (2009). Interactive Data to Improve Financial Reporting. *Release No. 33-8529*. Retrieved from https://www.sec.gov/rules/final/2009/33-9002.pdf.

Sepúlveda, A., Flores, D., & Sakuma, I. (2012). XBRL In Latin America. Available at: http://www.xbrlamericas.info/descargas/iBRNov12-LatAm.pdf .

Shan, Y.G., & Troshani, I. (2016). The effect of mandatory XBRL and IFRS adoption on audit fees : Evidence from the Shanghai Stock Exchange. *International Journal of Managerial Finance*, *12* (2), 109-135.

Shanmuganathan, M. (2016). Influences of XBRL & IFRS on Business Reporting – Canada. *Future Technologies Conference, San Francisco, United States.*

Shima, K. M., Yang, D. C., & Boulevard, C. B. (2012). Factors affecting the adoption of IFRS. *International Journal Of Business*, *17*(3), 277-298.

Shin, R.Y. (2003). XBRL, financial reporting, and auditing, *The CPA Journal*. Retrieved from http://archives.cpajournal.com/2003/1203/dept/d126103.htm.

Shkurti, R., & Allko, D. (2016). Business reporting language - A survey with the Albanian companies and institutions, *Management Strategies Journal*, *32*(2), 50-61.

Sinnett, W. M., & Willis, M. (2009). The time is right for standard business reporting. *Financial Executive*, *25*(9), 23–27.

Steenkamp, L. P., & Nel, G. F. (2012). The adoption of XBRL in South Africa: an empirical study. *The Electronic Library*, *30*(3), 409–425.

Sudalaimuthu, S., & Haraiharan, R. (2011). The role of XBRL and IFRS in financial reporting. *International Journal of Multidisciplinary Research*, *1*(4), 89-106.

Talebnia, G. (2016). XBRL for information quality and reducing information asymmetry, 383–390. *International Journal of Accounting and Financial Management*, *26*, 383-390.

Tan, C., & Shon, J. 2009. XBRL and its financial reporting benefits: Capital market evidence. *Working Paper, Fordham University, New York*.

Taylor, E. Z., & Dzuranin, A. C. (2010). Interactive financial reporting: An introduction to extensible business reporting language (XBRL). *Issues in Accounting Education, 25*(1)*,* 71-83.

Trites, G. (2010). The G-20 countries and their use of XBRL. Retrieved from http://www.xbrl.ca/docs/Seminar-April-7-2011/The_G-20_Countries_Use_of_XBRL.pdf.

Troshani, I., & Doolin, B. (2005). Drivers and inhibitors impacting technology adoption : A qualitative investigation into the Australian experience with XBRL. *18th Bled eConference eIntegration in Action Bled,* 1–16.

Troshani, I., & Doolin, B. (2007). Innovation diffusion: a stakeholder and social network view, *European Journal of Innovation Management*, *10* (2), 176-200.

Turki, H., Wali, S., & Boujelbene, Y. (2016). The effect of IFRS mandatory adoption on the information asymmetry. *Cogent Business & Management*, *3*(1), 1–19.

Vasal, V., & Srivastava, R. P.(2002). eXtensible Business Reporting Language (XBRL) – The Digital language of business: An Indian Perspective, *Indian Accounting Review*, *6*(1), 41-59.

Wada, Y. (2006). XBRL Case Studies: Showing XBRL in Use, The Bank of Japan. Retrieved from http://www.xbrl.org/CaseStudies/BoJ_XBRL_06.pdf .

Wallace, A. (2001). The new language of financial reporting. *Balance Sheet*, *9*(2), 29–32.

Wang, T., Wen, C. Y., & Seng, J. L. (2014). The association between the mandatory adoption of XBRL and the performance of listed state-owned enterprises and non-state-owned enterprises in China. *Information and Management*, *51*(3), 336–346.

Wang, Z., & Gao, S. S. (2012). Are XBRL-based financial reports better than non-XBRL reports? A quality assessment. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, *6*(4), 511–516.

Warren, M. (2004). Farmers online: drivers and impediments in adoption of Internet in UK agricultural businesses. *Journal of Small Business and Enterprise Development, 11*(3), 371 - 381.

Watson, L. A. (2010). Le langage mondial du gouvernement .2.0. *24 e conférence Internationale annuelle*. Retrieved from http://fr.slideshare.net/icgfmconference/liv-watson-icgfm-xbrl-a-language-of-the-government-world-francais.

Willis, M., Tesnière, B., & Jones, A. (2002). Corporate communications for the 21st Century. A white paper discussing the impact of Internet technologies on business reporting. Retrieved from http://www.xbrl.org/Business/General/PwC-Corp-Comms-21st-Century-2002-10.pdf.

WEF. (2011-2012). The Global Competitiveness Report 2011-2012. World Economic Forum.

Wunnava, P. V., & Leiter, D. B. (2009). Determinants of intercountry internet diffusion rates. *American Journal of Economics and Sociology*, *68*(2), 413–426.

XBRL France. (2006). XBRL: Introduction à la technologie XML des rapports financiers. Retrieved from http://www.nbb.be/DOC/BA/XBRL/Pub/Livre%20blanc%20XBRL%20France.pdf.

Xiao, J. Z., He, Y., & Chow, C. W. (2004). The determinants and characteristics of voluntary internet-based disclosures by listed Chinese companies. *Journal of Accounting and Public Policy*, *23*(3) ,191–225.

Yartey, C. (2008). The determinants of stock market development in emerging economies: Is South Africa different? *Working Paper*, International Monetary Fund (African

Department).

Yassin, M., El- Barghouthi, S., & Al-Khatib, I. (2012). XBRL: The future of financial reporting in Jordan. *International Conference on Business Intelligence and Knowledge Economy.*

Yoon, H., Zo, H., & Ciganek, A. P. (2011). Does XBRL adoption reduce information asymmetry ? *Journal of Business Research*, *64*(2), 157–163.

Yoshida, K. (2008). Tokyo Stock Exchange's XBRL Project, Tokyo Stock Exchange. Retrieved from http://archive.xbrl.org/18th/sites/18thconference.xbrl.org/files/Yoshida-TokyoStockExchangeXBRLProject.pdf.

Zeghal, D., & Mhedhbi, K. (2006). An analysis of the factors affecting the adoption of international accounting standards by developing countries, *The International Journal of Accounting*, *41*(4), 373–386.

Zeghal, D., & Mhedhbi, K. (2012). Analyzing the effect of using international accounting standards on the development of emerging capital markets. *International Journal of Accounting & Information Management*, *20*(3), 220–237.

Zehri, F., & Chouaibi, J. (2013). Adoption determinants of the International Accounting Standards IAS / IFRS by the developing countries. *Journal of Economics, Finance and Administrative Science*, *18*, 56–62.

Zhang, S. S., Riordan, R., & Weinhardt, C. (2013). Interactive data : Technology and cost of capital. *Accounting Information Systems for Decision Making*, 233–247.

Web Sites

www.xbrl.org

www.sca.gov.ae

www.xbrl.ca

www.mca.gov.in/xbrl

www.xbrl.dk

www.egov-estonia.eu/xbrl

http://www.mesiniaga.com.my/downloads/XBRL.pdf

Table 1: Presentation of 71 countries of the sample by region over seven –year period from 2009 to 2014 (see Appendix1)

Africa (10 countries)	Middle East and North Africa (10 countries)	America (12 countries)	Asia and Oceania (16 countries)	Europe (23 countries)
Egypt (0)	Israel (1)	Canada (1)	Japan (1)	Belgium (1)
South Africa (1)	Morocco (0)	United States (1)	South Korea (1)	Spain (1)
Botswana (0)	Bahrain (0)	Brazil (0)	Singapore (1)	France (1)
Cote d'Ivoire (0)	Jordan (0)	Chile (1)	Philippines (0)	United Kingdom (1)
Ghana (0)	Kuwait (0)	Mexico (0)	China (1)	Ireland (1)
Kenya (0)	Lebanon (0)	Peru (0)(1)	India (1)	Italy (1)
Mauritius (0)	Oman (0)	Ecuador (0)	Indonesia (0)(1)	Luxembourg (1)
Namibie (0)	Qatar (0)	Jamaica (0)	Malaysia (0)(1)	Netherlands (1)
Nigeria (0)	Tunisia (0)	Panama (0)	Bangladesh (0)	Portugal (0)
Zambia (0)	United Arab Emirate (UAE) (0) (1)	Trinidad & Tobago (0)	Kazakhstan (0)	Hungary (0)
		Argentina (0)	Pakistan (0)	Czech Republic (0)
		Colombia (0)	Slovakia (0)	Poland (1)
			Sri Lanka (0)	Russia (0)
			Vietnam (0)	Bulgaria (0)

New Zealand (0)	Croatia (0)
Australia (1)	Estonia (1)
	Latvia (0)
	Slovenia (0)
	Ukraine (0)
	Romania (0)
	Austria (0)
	Sweden (1)
	Switzerland (0)

The values of 0 and 1 are defined as follows:

1 if XBRL is adopted in the country (whatever the purpose of adoption : to meet the information needs of securities regulators or other government agencies);
0 it not

As Appendix 1shows, there are different years of adoption of XBRL for the same country (depending on the purpose of adoption: for securities regulators and other government agencies), we attribute the code 1 from the first adoption of XBRL. For example, in the case of United- States, we attribute the code 1 from 2005.

When we used the other measures for the adoption of XBRL:
For the adoption of XBRL to meet the information needs of securities regulators, the value 1 is attributed from the first year of use of XBRL in the country to meet the information needs of securities regulators.

For the adoption of XBRL to meet the information needs of other government agencies: if the country has different periods of adoption, we take into consideration the first adoption and we attribute the value 1 from the first year of using XBRL.

Some countries use XBRL as a project to assess the potential benefit that XBRL may offer. We didn't consider the use of project XBRL as adoption.

Table 2: Descriptive statistics of the independant variables

Variables	Obs	Mean	Std. Dev.	Min	Max
INVPRO	422	5.801	1.481	2.7	9.7
BGR	423	5.607	0.936	3	7.9
LED	426	1.374	3.820	-14.786	13.216
DEEO	426	4.699	13.815	-15.989	252.308

MES	417		3.990	3.506	-4.863	19.251
EDUC	421		43.656	25.438	2.3	103.1
TAB	422		5.074	0.628	3.6	6.5

		Number of country-years	Percentage of country-years
CULT	0	234	54.93
	1	192	45.07
ACCSYS	0	90	21.13
	1	336	78.87

INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; IEM: the importance of equity market; DEEO: the degree of external economic openness; MES : macro economic stability; EDUC : education; TAB: country's firm-level technology absorption; CULT: culture; ACCSYS : the accounting system (IFRS)

Table 3: Descriptive statistics of the dependant variable

Panel A: Summary statistics of XBRL for years 2009 -2014								
Variables	ADOPXBRL		ADOPXBRL-SR		ADOPXBRL-OGA		ADOPXBRL-SROGA	
	0	1	0	1	0	1	0	1
Frequence	295	131	354	72	333	93	392	34
Percent	69.25	30.75	83.10	16.90	78.17	21.83	92.02	7.98
Panel B: Yearly number of countries adopting XBRL from 2009 to 2014								
	0	1	0	1	0	1	0	1
2009	55	16	61	10	62	9	68	3
2010	52	19	60	11	58	13	66	5
2011	50	21	59	12	56	15	65	6
2012	47	24	59	12	53	18	65	6
2013	46	25	58	13	52	19	64	7
2014	45	26	57	14	52	19	64	7

ADOPXBRL: the adoption of XBRL to meet the information needs of securities regulators or one of the other government agencies; ADOPXBRL-SR: the adoption of XBRL to meet the information needs of securities regulators; ADOPXBRL-OGA: the adoption of XBRL to meet the information needs of other government agencies (which are: chamber of commerce, companies' house, statistical offices, tax administration, central balance sheet and banking regulators); ADOPXBRL-SROGA: the adoption of XBRL to meet the information needs of both securities regulators and at least one of other government agencies and zero otherwise.

Table 4: Comparison of variables between the two groups of countries adoption

Panel A: Based on the Student test			
Variable	Mean Variables		Sig.
	Adoption=1	Adoption=0	
LED	1.723	1.218	0.208

Panel B: Based on the Mann-Whitney test				
Variables	Rank Sum		Z	Sig.
	Adoption=1	Not adoption=0		
CULT	26990	63961	0.759	0.448
INVPRO	35742.5	53510.5	-6.961	0.000***
BGR	26191.5	63484.5	1.361	0.173
DEEO	27819	63132	0.127	0.899
MES	19966	67187	6.319	0.000***
EDUC	36229.5	52601.5	-7.629	0.000***
TAB	38180.5	51072.5	-9.046	0.000***
ACCSYS	28751.5	62199.5	-0.944	0.345

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; IEM: importance of equity market; DEEO: degree of external economic openness; MES: macro economic stability; EDUC: education; TAB: country's firm-level technology absorption ; ACCSYS: accounting system (IFRS)

***Significant at the 1% level

Table 5: Pearson correlation matrix

	ADOPXBRL	CULT	INVPRO	BGR	LED	DEED	MES	EDUC	TAB	ACCSYS
ADOPXBRL	1.000									
CULT	-0.031	1.000								
INVPRO	0.347 ***	0.297***	1.000							
BGR	-0.104 **	-0.274***	-0.135***	1.000						
LED	0.061	0.003	0.055	-0.101**						
DEED	0.155***	-0.061	-0.019	-0.090*	-0.025	1.000				

MES	-0.283***	0.263***	-0.112**	0.119**	0.165***	-0.057	1.000			
EDUC	0.368 ***	-0.374***	0.222***	0.150***	0.339***	-0.056	-0.455***	1.000		
TAB	0.444***	0.207***	0.217***	-0.369***	-0.115**	0.105**	-0.466***	0.273***	1.000	
ACCSYS	0.046	0.076	0.012	0.028	-0.104**	0.087*	-0.135***	0.147 ***	-0.034	1.000

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; IEM: the importance of equity market; DEEO: the degree of external economic openness; MES: macro economic stability; EDUC: education; TAB: country's firm-level technology absorption ; ACCSYS: accounting system (IFRS)

***Significant at the 1% level, **significant variable to 5% , *significant at the 10% level

Table 6: The pooled logitic regression analysis for the adoption of XBRL to meet the information needs of securities regulators or other government agencies (whatever the purpose of adoption)

Variables	Expected results	Coefficients	p-Value
CULT	+	-0.913	0.015**
INVPRO	?	0.503	0.000***
BGR	+	0.227	0.182
LED	+	0.157	0.004***
DEEO	+	0.052	0.010***
MES	+	0.027	0.662
EDUC	+	0.019	0.009***
TAB	+	2.054	0.000***
ACCSYS	+	0.479	0.200
Year Effect		Included	
Constant		-17.118	
R-squared		0.3454	

LR chi2	177.42
Prob > chi2	0.000
Number of observation	412

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; MES: macroeconomic stability; EDUC: education; TAB: country's firm-level technology absorption; ACCSYS: accounting system (IFRS); Year dummy (YEAR) represents dummy variables that reflect the years between 2009 and 2014.

***Significant at the 1% level, **significant variable to 5%.

Table 7: The pooled logitic regression analysis

		The adoption of XBRL to meet the information needs of securities regulators	The adoption of XBRL to meet the information needs of at least one of other government agencies	The adoption of XBRL to meet the information needs of both securities regulators and at least one of other government agencies
Variables	Expected results	Coefficients (p-Value)	Coefficients (p-Value)	Coefficients (p-Value)
CULT	+	-0.883	-0.031	0.516
		(0.033)**	(0.931)	(0.419)
INVPRO	?	0.570	0.011	-0.043
		(0.000)***	(0.912)	(0.796)
BGR	+	0.311	0.479	0.874
		(0.101)	(0.006)***	(0.005)***
LED	+	0.212	0.043	0.061
		(0.002)***	(0.432)	(0.545)
DEEO	+	-0.040	0.051	-0.096
		(0.159)	(0.005)***	(0.103)
MES	+	0.082	-0.038	0.065
		(0.283)	(0.553)	(0.542)
EDUC	+	0.015	0.031	0.051
		(0.077)*	(0.000)***	(0.000)***
TAB	+	2.047	1.388	1.674
		(0.000)***	(0.000)***	(0.002)***
ACCSYS	+	0.450	-0.614	-1.435
		(0.276)	0.106	(0.011)**

Year Effect	Included	Included	Included
Constant	-18.211	-13.232	-19.075
R-squared	0.3022	0.2565	0.3625
LR Chi2	114.42	112.87	85.11
Prob > chi2	0.000	0.000	0.000
Number of observation	412	412	412

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; IEM: the importance of equity market; DEEO: the degree of external economic openness; MES: macroeconomic stability; EDUC: education; TAB: country's firm-level technology absorption; ACCSYS: accounting system (IFRS); Year dummy (YEAR) represents dummy variables that reflect the years between 2009 and 2014.

***Significant at the 1% level, **significant variable to 5%, *significant variable to 10% level.

Table 8: Comparison between developed countries and developing countries –Pooled logistic regressions for the adoption of XBRL to meet the information needs of securities regulators or other government agencies (whatever the purpose of adoption)

		Sub-sample of developed countries	Sub-sample of developing countries
Variables	Expected results	Coefficients (p-Value)	Coefficients (p-Value)
CULT	+	0.552	-2.679
		(0.401)	(0.000)***
INVPRO	?	0.034	1.238
		(0.846)	(0.000)***
BGR	+	0.545	-0.405
		(0.085)*	(0.115)
LED	+	0.319	0.392
		(0.092)*	(0.000)***
DEEO	+	0.054	-0.175
		(0.053)*	(0.019)***
MES	+	0.364	0.209
		(0.115)	(0.021)**
EDUC	+	0.002	0.012
		(0.904)	(0.414)

TAB	+	-1.004	3.289
		(0.189)	(0.000)***
ACCSYS	+	-0.627	1.698
		(0.410)	(0.006)***
Year Effect		Included	Included
Constant		3.547	-24.569
R-squared		0.1771	0.4427
LR Chi2		23.97	116.48
Prob > chi2		0.0462	0.0000
Number of observation		114	298

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; IEM: the importance of equity market; DEEO: the degree of external economic openness; MES: macroeconomic stability; EDUC: education; TAB: country's firm-level technology absorption; ACCSYS: accounting system (IFRS); Year dummy (YEAR) represents dummy variables that reflect the years between 2009 and 2014.
***Significant at the 1% level, **significant at the 5% level, *significant at the 10% level

Table 9: Results of the logistic regression for the adoption of XBRL to meet the information needs of securities regulators or other government agencies (whatever the purpose of adoption)

Variables	Expected results	2009 (p-Value)	2010 (p-Value)	2011 (p-Value)	2012 (p-Value)	2013 (p-Value)	2014 (p-Value)
CULT	+	-2.649	-1.295	-0.750	-0.294	-0.634	-0.703
		(0.044)**	(0.250)	(0.450)	(0.751)	(0.500)	(0.441)
INVPRO	?	0.798	0.474	0.458	0.563	0.471	0.509
		(0.028)**	(0.131)	(0.116)	(0.031)**	(0.090)*	(0.049)**
BGR	+	-0.341	-0.360	0.711	0.276	0.478	0.552
		(0.504)	(0.422)	(0.191)	(0.575)	(0.306)	(0.200)
LED	+	0.170	0.258	0.089	0.110	0.220	0.425
		(0.114)	(0.115)	(0.564)	(0.435)	(0.242)	(0.090)*
DEEO	+	0.037	0.051	0.057	0.029	0.125	-0.033
		(0.435)	(0.224)	(0.343)	(0.469)	(0.146)	(0.622)

MES	+	0.067	0.099	0.006	-0.015	-0.015	0.114
		(0.639)	(0.544)	(0.974)	(0.943)	(0.933)	(0.495)
EDUC	+	0.009	0.040	0.021	0.028	0.022	0.023
		(0.695)	(0.051)*	(0.229)	(0.208)	(0.292)	(0.275)
TAB	+	2.765	1.757	2.406	1.628	2.149	2.321
		(0.009)***	(0.038)**	(0.006)***	(0.031)**	(0.011)**	(0.019)**
ACCSYS	+	0.453	0.795	1.397	-0.109	0.175	0.490
		(0.629)	(0.376)	(0.206)	(0.917)	(0.860)	(0.623)
Constant		-18.330	-14.336	-22.655	-15.305	-18.795	-20.723
R-squared		0.4025	0.3604	0.3728	0.3519	0.3745	0.3580
LR chi2		30.08	29.27	31.61	31.07	34.17	31.69
Prob > chi2		0.0004	0.0006	0.0002	0.0003	0.0001	0.0002
Number of observation		69	69	69	68	70	67

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; DEEO: the degree of external economic openness; MES : macroeconomic stability; EDUC : education; TAB: country's firm-level technology absorption; ACCSYS: accounting system (IFRS)

**significant variable to 5%, *significant variable to 10% level.

Table 10: Logit regression results for the adoption of XBRL to meet the information needs of securities regulators

Variables	Expected results	2009 (p-Value)	2010 (p-Value)	2011 (p-Value)	2012 (p-Value)	2013 (p-Value)	2014 (p-Value)
CULT	+	-3.287	-0.762	-0.989	-0.612	-0.821	-1.230
		(0.105)	(0.562)	(0.397)	(0.539)	(0.401)	(0.247)
INVPRO	?	1.068	0.543	0.630	0.493	0.391	0.607
		(0.031)**	(0.125)	(0.061)*	(0.106)	(0.215)	(0.053)*
BGR	+	-0.257	-0.165	0.702	0.406	1.009	0.634
		(0.713)	(0.759)	(0.247)	(0.453)	(0.084)*	(0.176)
LED	+	0.211	0.233	0.201	0.264	0.747	0.395
		(0.152)	(0.214)	(0.274)	(0.151)	(0.014)**	(0.213)
DEEO	+	-0.128	-0.056	-0.086	-0.060	0.013	-0.029
		(0.331)	(0.485)	(0.328)	(0.507)	(0.864)	(0.642)

MES	+	-0.083	-0.076	0.149	0.153	0.249	0.325
		(0.733)	(0.716)	(0.493)	(0.533)	(0.227)	(0.084)*
EDUC	+	-0.005	0.028	0.009	0.018	0.045	0.024
		(0.885)	(0.241)	(0.677)	(0.463)	(0.101)	(0.229)
TAB	+	3.087	1.363	2.891	2.429	3.156	1.996
		(0.045)**	(0.168)	(0.012)**	(0.019)**	(0.012)**	(0.079)*
ACCSYS	+	0.080	-0.074	0.559	0.808	0.873	0.871
		(0.940)	(0.939)	(0.605)	(0.496)	(0.467)	(0.451)
Constant		-20.998	-12.677	-26.210	-21.821	-31.221	-22.209
R-squared		0.4887	0.3500	0.3371	0.3140	0.3739	0.2701
LR chi2		27.90	21.19	21.49	19.90	25.12	17.80
Prob > chi2		0.0010	0.0118	0.0106	0.0185	0.0028	0.0375
Number of observation		69	69	69	68	70	67

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; IEM: the importance of equity market; DEEO: the degree of external economic openness; MES : macroeconomic stability; EDUC: education; TAB: country's firm-level technology absorption ; ACCSYS : accounting system (IFRS)

**significant variable to 5%, *significant variable to 10% level.

Table 11: Logit regression results for the adoption of XBRL to meet the information needs of at least one of other government agencies

Variables	Expected results	2009 (p-Value)	2010 (p-Value)	2011 (p-Value)	2012 (p-Value)	2013 (p-Value)	2014 (p-Value)
CULT	+	-0.453	-0.250	0.001	0.666	-0.079	0.144
		(0.644)	(0.830)	(0.999)	(0.462)	(0.929)	(0.877)
INVPRO	?	0.059	-0.185	-0.090	0.100	0.004	0.037
		(0.833)	(0.536)	(0.733)	(0.673)	(0.988)	(0.881)
BGR	+	0.315	0.450	0.616	0.638	0.862	0.489
		(0.518)	(0.390)	(0.229)	(0.224)	(0.072)*	(0.247)
LED	+	0.015	-0.008	-0.050	0.048	0.141	0.557
		(0.903)	(0.961)	(0.740)	(0.755)	(0.478)	(0.044)**
DEEO	+	0.064	0.079	0.054	0.028	0.081	-0.095

		(0.174)	(0.045)**	(0.342)	(0.209)	(0.146)	(0.124)
MES	+	-0.063	0.052	-0.003	-0.010	0.021	0.060
		(0.702)	(0.801)	(0.988)	(0.956)	(0.906)	(0.737)
EDUC	+	0.017	0.058	0.042	0.050	0.033	0.051
		(0.491)	(0.020)**	(0.037)**	(0.049)**	(0.117)	(0.039)**
TAB	+	1.425	1.663	1.656	1.021	1.927	1.440
		(0.163)	(0.074)*	(0.051)*	(0.192)	(0.022)**	(0.129)
ACCSYS	+	-0.218	-0.351	0.185	-1.808	-0.981	-0.847
		(0.827)	(0.715)	(0.862)	(0.100)*	(0.300)	(0.381)
Constant		-12.326	-14.984	-15.408	-12.145	-17.157	-14.345
R-squared		0.2207	0.3334	0.2963	0.2995	0.2755	0.2981
LR chi2		11.79	22.26	21.41	23.54	22.55	23.82
Prob > chi2		0.2254	0.0081	0.0110	0.0051	0.0073	0.0046
Number of observation		69	69	69	68	70	67

CULT: culture; INVPRO: strength of investor protection; BGR: burden of government regulation; LED: level of economic development; DEEO: the degree of external economic openness; MES: macroeconomic stability; EDUC: education; TAB: country's firm-level technology absorption ; ACCSYS : accounting system (IFRS)

**significant variable to 5%, *significant variable to 10% level

Appendix 1: XBRL adoption on the world

Countries	BR	CBS	SMR	BRG	CH	TA	SO	Sources
Argentina	* **P**							O'Kelly (2010)
Australia	* (2010) **V**	* (2010) **V**	* (2010) **V**	* (2010) **V**		* (2010) **V**	* (2010) **V**	website of XBRL international organization (2013) (www.xbrl.org)

								O'Kelly (2010) Debreceny, Farewell, and Verkruijsse (2012) Chen (2012) Trites (2010)
Belgium	* (2007)**M**	* (2007)**M**				* (2011)	* (2009)	Dufresne (2010) De Brandt (2012)
Brazil	* **P**					* **P**		Ernst and Young (2010)
Canada			* (2007) **V**					www.xbrl.ca
Chile			* (2009) **V** (2010)**M**					O'Kelly (2010) IASC Foundation XBRL Team (2010) Sepúlveda, Flores, and Sakuma (2012)
China			* (2004)**M**					Peng, Shon, and Tan (2011)
Colombia	* **P**							O'Kelly (2010)
Cyprus	* **M**							Maguet (2012)
Estonia				* (2010 **V**) (2012**M**)				www.egov-estonia.eu/xbrl Maguet (2012)
Finland	* **V**							Maguet (2012)
France	* (2007)**M**			* (2009)**M**				Maguet (2006) Pasmooiji (2010) Ernst and Young (2010) Trites (2010) Maguet (2012)
Greece	* **M**							Pasmooiji (2010)

Hong Kong			* **V**					website of XBRL international organization (2013) (www.xbrl.org) Pasmooiji (2010)
India	* (2008)**M**		* (2011)**M**					O'Kelly (2010) www.mca.gov.in/xbrl Trites (2010)
Indonesia	* (2012)**M**							Rahwani (2013)
Ireland	* **V**			* (2011) **P**		* (2012 **V**) (2013**M**)		website of XBRL international organization (2013) (www.xbrl.org) Pasmooiji (2010) Maguet (2012) PWC (2013)
Israel			* (2008)**M**					O'Kelly (2010) Markelevich, Shaw, and Weihs (2015)
Italy				* (2009)**M**				O'Kelly (2010) Pasmooiji (2010)
Japan	* (2006)**M**		* (2008)**M**			* (2004)**M**		Sakuta (2008) Wada (2006) Yoshida (2008) O'Kelly (2010)
Lithuania	* **M**							Pasmooiji (2010) Maguet (2012)
Luxembourg	* (2008)**M**			* **V**			* **V**	O'Kelly (2010) Pasmooiji (2010) Maguet (2012)
Malaysia	* (2012)**M**							http://www.mesiniaga.com.my/downloads/XBRL.pdf Francis (2012)
Mexico	* **P**							O'Kelly (2010)

Netherlands	* (2011) **V**	* (2011) **V**		* (2011) **V**		* (2011) **V**	* (2011) **V**	website of XBRL international organization (2013) O’Kelly (2010) (www.xbrl.org) Debreceny, Farewell, and Verkruijsse (2012) Chen (2012)
New Zealand						* **P**		website of XBRL international organization (2013) (www.xbrl.org) O’Kelly (2010)
Peru	* (2013) **V**		* (2013) **V**					Sepúlveda, Flores, and Sakuma (2012)
Poland	* **M**		* (2011) **P**		* (2010)**M**			website of XBRL international organization (2013) (www.xbrl.org) O’Kelly (2010) Pasmooiji (2010)
Russia	* **P**							website of XBRL international organization (www.xbrl.org)
South Africa			* (2009) **v**					O’Kelly (2010) Debreceny, Farewell, and Verkruijsse (2012)
South Korea	* (2010) **V**		* (2007)**M**					O’Kelly (2010) Yoon, Zo, and Ciganek (2011)
Spain	* (2008)**M**	* (2007) **V**	* (2005)**M**	* (2009)**M**				O’Kelly (2010)
Sweden				* (2006)**V**				Debreceny, Farewell, and Verkruijsse (2012)
Switzerland	*							O’Kelly (2010)

	P							
Taiwan			* (2010)**M**					website of XBRL international organization (2013)(www.xbrl.org)
Thailand			* **V**					website of XBRL international organization (2013)
Turkey			*					website of XBRL international organization (2013) (www.xbrl.org)
United Arab Emirates			* (2011) **P** (2014) **M**					www.sca.gov.ae website of XBRL international organization (www.xbrl.org)
United States	* (2005)**M**		* (2005 **V**) 3 phases (2009)**M** (2010)**M** (2011)**M**					FFIEC (2006) SEC (2009) O'Kelly (2010) PWC (2013)
United Kingdom					* **V** (2013**M**)	* (2011)**M**		O'Kelly (2010) Debreceny, Farewell, and Verkruijsse (2012) Maguet (2012)

XBRL is adopted by companies in a country to meet the information needs of :

BR: Banking Regulator ; **CBS:** Central Balance Sheet ; **SMR:** Stock Market Regulator ; **BRG:** Business Register ; **CH:** Companies House ; **TA:** Tax Administration ; **SO:** Statistical Offices

V: voluntary adoption of XBRL ; **M:** mandatory adoption of XBRL ; **P:** the adoption of XBRL as a pilot project

The year of adoption of XBRL is indicated in brackets

We did not include in our sample the countries in which we do not have enough information regarding the adoption of XBRL.

DECISION SCIENCES INSTITUTE

A Better Approach for Mishandled Bags Industry' Measure

Ana Caroline Costa
Ana.costa07@latam.com
Embry Riddle Aeronautical University
Daniel B Ravagnani
daniel.ravagnani@smiles.com.br
Embry Riddle Aeronautical University
Domenico Giuseppe Valente
d.valente7@gmail.com
Embry Riddle Aeronautical University
Vitor Coletti Fernandes
vitorcofer@gmail.com
Embry Riddle Aeronautical University
Dr. Leila Halawi
Embry Riddle Aeronautical University

ABSTRACT

Since 1987, the U.S. Department of Transportation (DOT) has ordered U.S. airlines to report baggage handling statistics. The Code of Federal Regulations (CFR) requests each U.S. carrier to report monthly the total number of passengers enplaned and the total number of mishandled baggage reports (MBR), including lost, stolen, damaged, and delayed baggage. The U.S. DOT also created a mishandled baggage rate, representing the total number of MBRs divided by 1,000 passengers enplaned. This KPI is monthly published at the Air Travel Consumer Report. And it is used worldwide to compare regions, airports, and airlines, characterizing efficiency and consistency in service and setting a benchmark for the industry. The problem with this indicator created in 1987 is that it only considers passengers, not checked baggage. The customer's behavior has changed a lot after airlines started charging for the checked baggage, influencing KPI results. Therefore, in January 2019, US DOT proposed a new mishandled bags KPI for the airline industry; instead of dividing the original KPI per 1,000 passengers, the new KPI proposes dividing the total number of mishandled baggage reports per 1,000 checked baggage. This KPI considers checked bags, corrects this flaw, and creates a more accurate and reliable indicator, avoiding incorrect investment decisions and losing money.

Keywords: U.S airlines, KPI, mishandled baggage

INTRODUCTION

According to Peter Drucker (2009), one of the men responsible for modern business management, "if you can't measure it, you can't improve it.". This well-known statement sets the importance of having accurate and reliable data and models of measurements to help companies solve problems and minimize or avoid wrong decisions. This concept also applies to the airline's operations. The procedures performed must be measured and monitored to allow companies to improve their ability of decision-making. Airports and airlines are doing their best to achieve higher levels of customer satisfaction to translate into continuously increasing demand for their services. Both airlines and airport managers use mathematical data through metrics and indicators, like other industries, to problem-solve and structure business decisions before taking action. That is the main reason why ensure that the measurements are correct is so crucial for any company. The aviation industry is dynamic and passes every day for constant changes, and indicators also need to keep pace with these changes.

Linked to customer satisfaction, excellence in the baggage management process is crucial and has an important rule to define the overall customer experience through airlines and airports. According to the ANAC Consumer Monitoring Report - 1st quarter of 2019 (ANAC,2019), baggage is the fifth in the ranking of reasons why customers are complaining, with 12.3% of the total number.

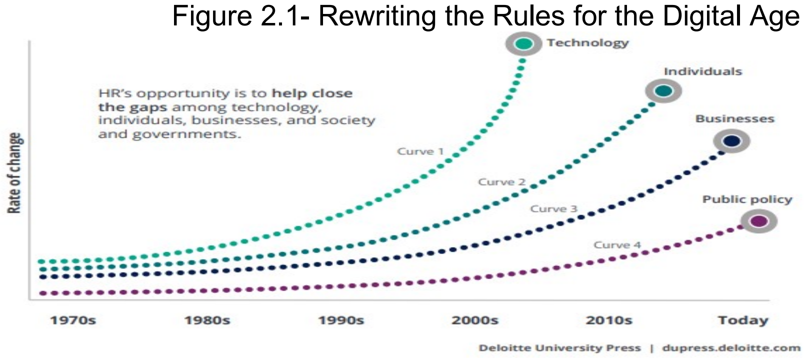
This study aims to analyze whether the current use of U.S. DOT's KPI for measuring the mishandled baggage performance meets the airline industry's needs. Also, this study evaluates this new and more robust indicator that might bring financial benefits to the industry. The researchers strongly believe a stronger correlation between mishandled bags and checked bags than mishandled bags and passengers. The study analyzes if, instead of using the number of carried passengers, KPI's mishandled bags were calculated using mishandled bags and divided by 1,000 checked baggage.

Review of the Relevant Literature

The economy, customers, habits, and technology have been changed faster day by day. This evolution is faster in the airline industry because of the dependency of both customers and technology. To keep up to date with this rapid evolution and natural trend of the market and passengers, airlines have to adopt new measures and processes (Forbes, 2018).

A study from Deloitte Press University (2017) shows the speed of changes in various aspects of our society. The image below (Figure 2.1) shows us that technology has been changing much faster than any other factor included. This change creates a gap of the curves and turns individuals, businesses, and public policy obsolete quickly.

Figure 2.1- Rewriting the Rules for the Digital Age



So, it is profusely clear that technology is advancing at an unprecedented rate. The business's productivity could not keep the same pace and might be one reason companies are turning disrupted faster (Deloitte University Press, 2017).

According to Daniel Power (2013), managers are forced to make quick decisions in a complex, inconstant, and uncertain environment. Most managers want and need more and better analyses and decision-relevant reports to support the decision-making process to stay forward in this context. According to Wayne (2006), "KPI is a metric to measure how well the organization or an individual performs an operational, tactical or strategic activity that is critical for the current and future success of the organization." The KPIs began taking place in the world, supporting the decision-making in the companies. KPIs are relevant to business goals because they keep them steps ahead of decisions. (Poleski, D. 2019)

But it is essential to keep in mind that choosing the correct KPIs and metrics is critical because it will reflect better results and decisions. The best KPIs aren't standard metrics. One crucial direction expected from the KPIs is to help companies predict future actions and understand the best choice. When the KPIs are well developed and provide accurate and up-to-date information, it is possible to understand, for example, the entire experience of the customer from the beginning to the end (Barlow, 2015).

To discover new opportunities, people need to go deeper to find the metrics representing what is going on (Barlow, 2015). Therefore, it is necessary to keep KPIs updated and reflecting the reality of the company. One wrong decision driven by the wrong KPI can affect the company.

Key Performance Indicator (KPI) and the Airline Industry

The Key Performance Indicator in the airline industry became more relevant after the airline deregulation act in 1978 (McDermott, 2017). The U.S. government maintained the statistics on numerous metrics of airline service quality (Rhoades & Waguespack, 2008). However, these reports became accessible to the public with the Air Travel Consumer Report (ATCR) in 1988.
The ATCR discloses the Mishandled Baggage Rate KPI, which the airline industry uses to analyze the performance of luggage processes (Rhoades & Waguespack, 2008). Many companies, organizations, and researchers have adopted the Air Travel Consumer Reports information to categorize a service quality or create new metrics. For example, the Airline Quality Rating (AQR) considers 19 quality factors and includes Mishandled Baggage (Bowen, Headley, & Luedtke, 1992).

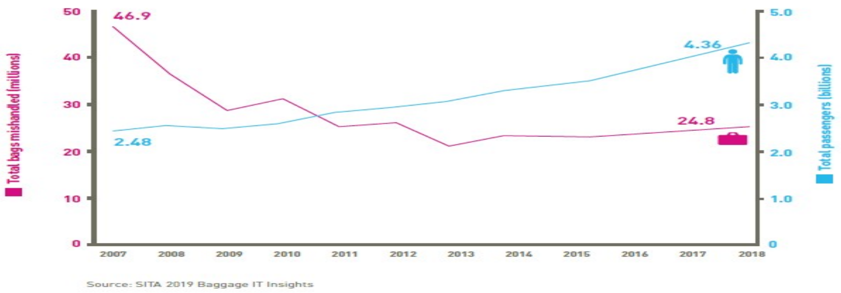
The History of Mishandled Baggage Rate

The airlines' mishandled baggage reports began in October 1987, when the U.S. Department of Transportation (DOT) has requested the U.S. carriers to report every month the number of transported passengers system-wide and the mishandled-baggage occurrences, including lost, pilfered, damaged, and delayed baggage (Gov_info, 14 CFR 234, 2017). The U.S. DOT also created the mishandled baggage KPI, the total number of MBRs divided by thousands of passengers enplaned. This number is monthly published in the Air Travel Consumer Report (ATCR).

Annual Surveys and Reports Published with MBR to the Airline Industry

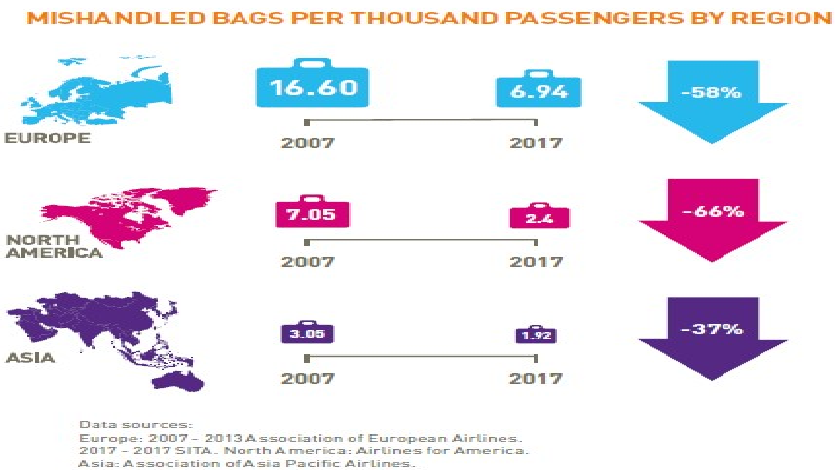
Several reports disclose the MRB, and Figure 2.2 below illustrates one of them: SITA's annual report. This image highlights that from 2007 to 2018, the number of baggage has dropped from 46.9 million to 24.8 million. On the other side, the number of passengers has increased from 2.48 billion to 4.36 billion.

Figure 2.2- Long Term Decrease in Baggage Mishandling



The following image (Figure 2.3) shows the variation in the mishandled baggage rate. Over the ten years, the KPI suffered a significant reduction in different world countries, emphasizing that, especially in North America, there was a reduction of 66% (SITA, 2019).

Figure 2.3- Baggage Report



Therefore, the researchers have seen a significant increase in passengers in recent years and reduced mishandled luggage KPI. This scenario is attributed to investments in airport procedures and improved infrastructure (SITA, 2019). However, it is not until 2018 that IATA makes one official procedure guide for bag tracking (IATA, 2018).

Inconsistencies, Errors, or Constraints identified to MBR

In 2001, the Bureau of Transportation Statistics (BTS) published the Mishandled Baggage Report, excluding non-revenue passenger numbers. Because of that, BTS highlighted that the KPI did not accurately represent the carrier's capability to handle baggage, being inappropriate for cross-carriers analysis. She also questioned that the reports do not express the headaches consumers face in solving their luggage problems such as late arrival, damage, or stolen.

The 14 CFR 234 and 250 (2017) reports data threshold for mishandled baggage is one percent of the Industry Scheduled-Service Domestic Passenger Revenues. The report is not intended to consider whether few passengers are avoiding checking baggage to avoid fees. In 2018, Charlie Leocha (Leocha,2017), president of Travelers United, wrote a distortion in Mishandled Baggage KPI. According to her, the reason is that airlines use to divide the total occurrences of baggage in numbers per 1,000 carried passengers, instead of dividing by the number of checked bags. She reports the case of two airlines that have the same baggage handling procedure, but one seems to be much better than the other because its passengers check fewer bags.

The Influence of New Commercial Aviation Market on the MBR

The theme of checked baggage has become increasingly relevant due to the growing concern about flight safety. Since the unfortunate event of 9/11, political, economic, and social factors have spawned a huge crisis in the commercial aviation segment, as passenger numbers have declined sharply.

The scenario of severe economic recession has forced airlines to adapt financially and economically to the low passenger demand. Due to this complex situation, the industry began to show a high baggage breakdown and loss rate. This situation lasted for six years until mid-2007. All of this was only made evident in 2012 in a letter sent to the U.S. House of
Representatives by the Committee on Trade, Science and Transport of the Government Accountability Office (GAO). This letter contained an analysis of DOT information that showed an apparent deterioration in Baggage Claim Rate (GAO-12-804R; P.6)

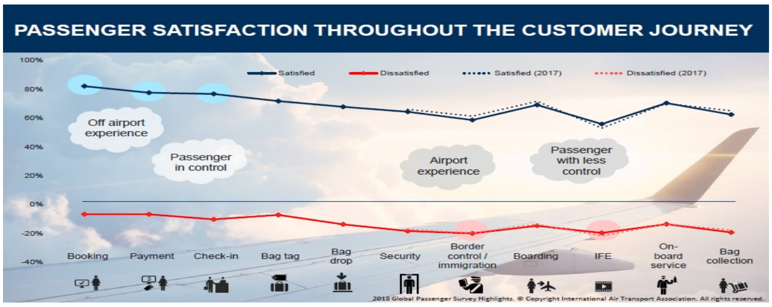
On March, 21st 2017, the U.S. Department of Transportation effects a new policy(www.govinfo.gov,14 CFR Parts 234 e 241), requiring airlines to report the number of baggage checked on each flight. However, most airlines in the world, especially the Brazilian ones, continue to use the existing baggage KPI to reference all their internal analysis and decision-making.

Baggage Fee and Airline Industry

In May 2008, there was a major milestone that altered the worsening trend in the Mishandled Baggage KPI. American Airlines, the world's largest airline, has started charging $15 for many passengers to check their first baggage in the United States (NYtimes, Maynard, 2008). After that, most airlines, including Delta and United, have followed that decision and created similar fees. Europe's Lowest Cost airline instituted baggage charges on its flights in 2006. It charges about $32 for each bag checked at the airport. Also, airlines like Ryanair strongly discouraged checked bags on their flights, as stated by his CEO, Michael O'Leary, in a Wall Street Journal interview (quoted in Allon, 2009). He commented that payment for checked baggage is not related to revenue. Instead, he had indicated that passengers need to change travel behavior and avoid checking baggage, helping carriers to reduce the airport handling costs.

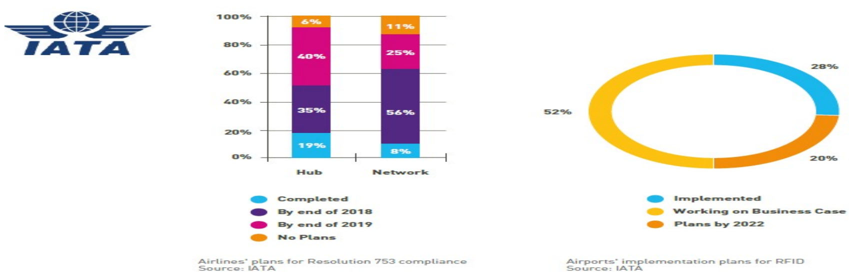
Passenger behaviors and preferences have changed in the aviation industry over the past years. IATA has made annual survey reports since 2012 to identify the behaviors and preferences of the passengers. Figure 2.4 shows the result of the 2018 survey with more than 10,400 respondents across 150 countries that shared their travel preferences.

Figure 2.4- IATA 2018 Global Passenger Survey Highlights



According to the SITA 2019 Baggage-IT-Insight 2019, eight out of ten passengers check-in their luggage, most traveling with one bag. In 2018 was around 4.3 billion bags were carried. With IATA Resolution 753 on baggage tracking in effect, the industry is looking at RFID as a low-cost solution. Figure 2.5 shows the completion rate of airlines and airport members of IATA to comply with Resolution 753.

Figure 2.5- SITA Baggage-it-insight



The Importance of a Good KPI for the Right Investments

A good KPI should convert an airline's business strategy into manageable operations based on the Data you monitor. After all, the old cliché: "What gets measured gets done" (Drucker, 1954) always rings true. KPIs are used to understand and improve a company's progress. To make the right decision, drive business investments, invest in data compilation, accurate business, and leave the usual behind-old feelings and assumptions.

Airlines' main argument for implementing baggage fees would be to provide lower-endures airfares generated by lower operating costs, reduced check-in staff and baggage handlers, and lower fuel consumption. Another significant consequence would be less luggage transferred between flights and increased work agility for baggage handlers. The biggest benefit of this would be for customers as flights depart from the ground at scheduled departure times.

According to the 2019 SITA Baggage I.T. Insights report, the total number of handled bags fell by approximately 90% in 11 years from 2007 to 2018, even with the large growth in total transported passengers in the period. This drop is due to passengers who have less baggage checked and the high investments in technology in the sector, especially in baggage tracking. Of the total checked baggage, 0.6% is misdirected or lost each year; more than half of this baggage is lost due to problems with flight transfers. The other half of the lost luggage is due to flight delays or mishandled by the baggage operators (SITA, 2019).
Generally, these baggage operators are independent third-party employees who provide services to various airlines or are directly hired by the airlines. As such, the quality of this service can vary dramatically. Several corporations worldwide found a strong relationship between NPS score and company growth rate (Reichheld's, 2011) compared to the score of their largest competitors in the market. Philips' annual report provided updates to the NPS reviews and announced that 60% of the company's revenue came from business in leadership or leadership positions in NPS. In this sense, remodeling the mishandled baggage KPI, according to our proposal, would generate major financial and non-financial benefits for customers and airlines.

The Importance of Baggage for Passenger Experience

According to IATA Business Care Report data, 22.7 million mishandled luggage in 2018, a rate of 5.57 per 1,000 passengers. Although this figure equated to a 10% decrease over the previous year, the cost to the airline industry was $ 2.3bn. If the airline does not deliver baggage in Brazil, the Passenger may require financial compensation to purchase necessities. For instance, if the baggage is not found by the carrier and not returned within seven days in domestic and 21 days for international flights, then the Passenger is entitled to request baggage compensation.

Also, the Passenger may take legal action for compensation for moral and material damages due to loss of luggage, which is very common in Brazil. Passengers highly perceive the quality of airports and airlines through baggage handling. Passenger satisfaction and the performance of baggage operations are important aspects of the success of airport operations and are closely interconnected.

This is especially true following resolution 753 from IATA, which encourages the carriers to reduce mishandling bags through a cross-industry tracking system following the baggage journey (SITA, 2019). There are many efforts by airports and carriers to perform baggage tracking and improve their luggage processes. The report IATA Business Care Report data (December 2018) says that end-users perception of baggage operations has been changing. When surveyed, 84% of passengers stated that they would like to have bag tracking; two-thirds expect this functionality, and the remaining believe that the additional services will be beneficial.

With the increasing use of baggage tracking data by Airlines and airports, RFID tagging also keeps pace with this evolution. Radio Frequency Identification (RFID) technology captures Data using a radio frequency. RFID provides more valuable information about a baggage journey and will help make the process more efficient. Using artificial intelligence (A.I.) in baggage handling will help enhance the passenger experience, improve operations, and reduce costs. "Few essential truths endure in time - but one of them is the notion that we should strive to turn our clients into enthusiastic advocates who comment wonderful things to our friends and colleagues." (Reichheld, 2011).

Use of KPI as a reference in performance contracts with third parties

Because it is a more operational function and does not represent the core of aviation, many airlines outsource ground handling. This number is around 50% worldwide (IATA, 2018). By providing checked baggage KPI-based solutions with innovative digital analytics applications, airlines will monitor third-party performance better, suggest improvements in baggage journey processes, and reduce costs with contracts that continue to charge airports and airlines for boarded passengers. As customer behavior changes to check for less baggage, third parties tend to use fewer human resources, leading to a lower cost suggestion (IATA, 2018).

Accurate KPI is Important and Contributes Directly to the Result

The KPIs are important because they offer a business diagnosis to a better position in the market. Consistent KPIs and metrics must be connected to the business purpose to achieve the desired success. "If you don't know where you want to go, then it doesn't matter which path you take." (Carroll, 1971)

Key performance indicators are important methods that companies use to measure the potential to achieve their goals (Durcevic, 2019). Also, KPIs could help airlines to obtain more effective key performance variables. Factors such as net income and EBITDA are often the main triggers of the company's well-known recognition and profit-sharing programs and revenue, operating efficiency.

There are other opportunities for airline companies with KPI utilization, which could indirectly impact customer satisfaction and employee work performance. These opportunities help improve attitudes and generate quality services. Companies that use KPIs for this purpose can produce more consistent results.

In the aviation market, keeping the customer satisfied and happy based on good services and processes became the company's success and a market strategy to build customer loyalty. A satisfied and happy customer makes new purchases or services and promotes the company to family and friends. Having robust and accurate KPIs is of utmost importance to any organization obstinate in growing and profiting.

Being aware of your customers' level of satisfaction, constantly analyzing your market positioning, improving your production, sales, and processes are key to thriving. Successful implementation of KPIs within a business brings time optimization and cost savings. The investments required to implement KPIs include new technologies, staff qualification, and process improvement primarily. But this cost will be rewarded with competitiveness gains.

Methodology

This study compares the actual Mishandled Baggage KPI and the proposed one, using the checked bags. The researchers applied linear regression to compare those two indicators and define which of the two independent variables, passengers or checked bags, has the highest correlation with the dependent variable, the number of mishandled bags.

The following data was gathered from the three airlines: (1) the number of mishandled bags, (2) the number of checked bags, and (3) the number of passengers carried. All database transformations, calculations, and statistical analyses were developed with the help of Excel multifunction software, a program from Microsoft Corporation. The data layout was also elaborated graphically to aid the analysis and facilitate their interrelation process. The researchers also performed a Correlation and Simple Linear Regression analysis. This technique shows which independent variables, the number of checked bags or the number of passengers carried, have the highest correlation with the dependent variable, the number of mishandled bags.

In addition, the researchers analyzed a real case scenario, comparing data extracted from one Brazilian airline for two different routes, and demonstrated how the change in the denominator dividing by baggage would change the key performance results.

The Simple Linear Regression Model (SLRM)

According to Kazmier (1982, p.299), linear regression analysis is used to forecast the value of a dependent variable since the value of one or more associated independent variables is known. The regression equation is the algebraic formula used to predict the value of the dependent variable.

However, before estimating the simple linear regression model (SLRM), one must consider other deterministic factors influencing the model parameters. This study addresses two other essential points for analysis, Pearson's Linear Correlation Coefficient and The Determination coefficient ($R^{!}$). According to Melo (2012), Pearson's Correlation Coefficient ($\rho(X; Y)$) aims to measure the degree of association between two variables whose value is between -1 and 1. When X and Y vary in the same direction, the correlation is positive, so $\rho > 0$, if ◌̊ X and ◌̣ Y vary; otherwise, the correlation is negative, so $\rho < 0$. If $\rho = 1$, the correlation is perfect positive, If $\rho \neq 1$, the correlation is perfect negative, If $\rho = 0$, the correlation is null. Correlation Ranges are weak from 0 to |0.5|; reasonable from |0.5001| to |0,7|; and strong from |0.7001| to |1| (Melo, 2012).

The coefficient of determination, $R^!$, is interpreted as the proportion of Y's observed variability, explained by the model considered. The value of $R^!$ It belongs to the range[0; 1], and when closer to 1, the better the model fit.

Outcomes

To observe the correlation of the dependent variable, mishandled bags, with the independent variables, the researchers plotted the two graphs below, weekly mishandled bags per number of passengers (Figure 4.1) and weekly mishandled bags per checked bags (Figure 4.2). Also, they drew the trend line with the linear equation and the coefficient of determination for each of the images.

Figure 4.1- Weekly Mishandled Bags Per Number of Passengers

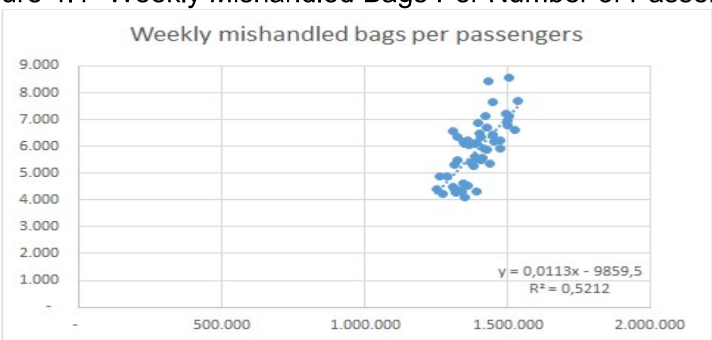
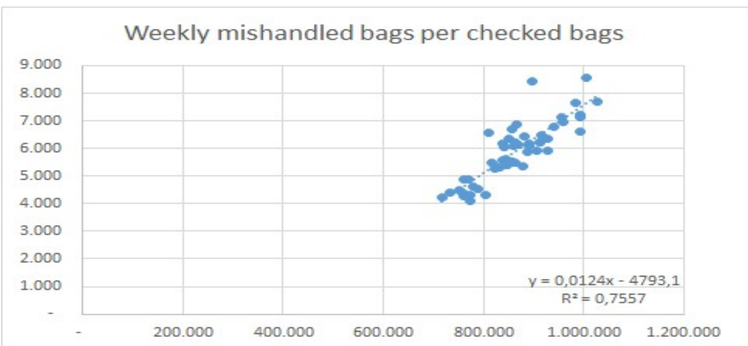


Figure 4.2- Weekly Mishandled Bags Per Checked Bags



A larger the $R^!$ The model explains the better de relationship between the two variables and that the trend line has a better fit to the numbers in the sample. As seen in the charts, the current KPI model used as the denominator the number of passengers has $R^{!}$ of 52.12%. Therefore, in 48% of cases, the dependent variable cannot be explained by the regressors present in the current model. However, the coefficient of determination is 75.57% in the weekly mishandled bags per checked bags' scenario, showing an increase of 23 percentage points. This 45% increase in $R^{!}$value indicates that the number of lost bags is much better explained by the number of checked bags, favoring the change in how the airlines calculate the KPI today.

The researchers performed the linear regression on the same data set (week performance). They used the value of Mishandled Baggage as the dependent variable and the other two variables, number of passengers and number of checked bags, as independent variables.

The result of this linear regression analysis shows that the Checked Baggage variable has a much lower Pearson's Linear Correlation Coefficient (P-value). And, as explained in the methodology chapter, a lower P-value shows a higher degree of association between two variables.
The result proves that the Checked Bags variable has a better correlation than the Mishandled Baggage's occurrences. Table 4.1 shows the results for the Simple Linear Regression Model analysis.

Table 4.1- Simple Linear Regression Model Results for Checked Bags and Passengers as Independent Variables and Mishandled Baggage Occurrences as the Dependent Variable

SLRM	**Coefficients**	**Standard error**	**Stat t**	**P-value**
Intersection	-1154,770257	1873,157469	-0,616483278	$5,4043*10^{(\$}$
Checked Bags	0,01673507	0,002231895	7,498145245	$1,1266*10^{()}$
Passengers	-0,005313407	0,002452692	-2,166357501	$3,5176*10^{(!}$

Note: P-value wrote in scientific notation. The Regression Adjusted R Squared is 0,7679.

After, the researchers performed two linear regressions independently, considering only one independent variable in each one. As expected, the P-value remains lower for the Checked Baggage variable than the p-Value of the Passenger Transported variable in the same period.

The tables below, Tables 4.2 and 4.3, show the result of these regressions.

Table 4.2

Simple Linear Regression Model results for Passengers as the independent variable and

Mishandled Baggage occurrences as the dependent variable

SLRM	Coefficients	Standard error	Stat t	P-value
Intersection	-9859,55	2132,521	-4,62342	$2{,}69*10^{(*}$
Passengers	0,01129	0,00153	7,377654	$1{,}54*10^{()}$

Note: Adjusted R Squared is 0,5116.

Table 4.3

Simple Linear Regression Model results for Checked Bags as the independent variable and

Mishandled Baggage occurrences as the dependent variable

SLRM	Coefficients	Standard error	Stat t	P-value
Intersection	-4793,14	859,5704	-5,5762	$9{,}89*10^{(+}$
Checked Bags	0,01237	0,000995	12,43586	$6{,}42*10^{(\$+}$

Note: Adjusted R Squared is 0,7508.

Better results in the current KPI but less checked baggage per Passenger

Analyzing the historical Data for Mishandled Baggage's Reports, the U.S. airlines were able to reduce in 2016 the mishandled baggage KPI to a rate of 2.70 per 1,000 passengers. The lowest annual rate since 1987, the year of the U.S. Department of Transportation (US DOT) began tracking results (Bureau of Transportation Statistics, 2017). The image below (Figure 4.3) shows the annual evolution of this indicator per Passenger.

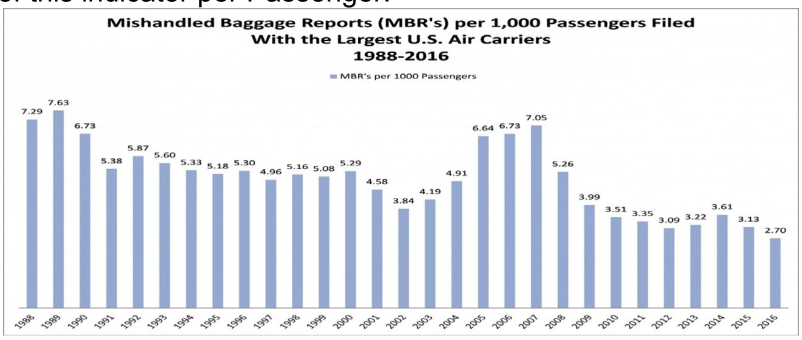


Figure 4.3 Shows the mishandled baggage reports (MBR) for the Largest U.S. Carriers from 1988 to 2016.

In October 2018, the Bureau of Transportations Statistics published an addendum Nº 30 Technical directive - updating the KPI to the total number of mishandled bags per the number of bags enplaned. As the addendum is only valid after January 2019, there is no historical data yet for the comparison. The table below results from the July 2019 KPI; the new KPI has a 6.33 rate while the "passenger KPI" shows 2.4 (2017 SITA report results for North America).

AIR TRAVEL CONSUMER REPORT

MISHANDLED BAGGAGE: RANKING OF U.S. REPORTING MARKETING CARRIERS (MONTHLY)

RANK	CARRIER*	JULY 2019		
		NUMBER OF BAGS ENPLANED	NUMBER OF BAGS MISHANDLED	NUMBER OF BAGS MISHANDLED PER 1000 ENPLANED
1	**ALLEGIANT AIR**	**808,594**	**1,614**	**2.00**
2	**FRONTIER AIRLINES**	**1,167,391**	**4,563**	**3.91**
3	**ALASKA AIRLINES NETWORK**	**2,862,127**	**12,294**	**4.30**
	- ALASKA AIRLINES	2,170,020	9,382	4.32
	- BRANDED CODESHARE PARTNERS	692,107	2,912	4.21
4	**HAWAIIAN AIRLINES NETWORK**	**618,019**	**2,751**	**4.45**
	- HAWAIIAN AIRLINES	601,288	2,513	4.18
	- BRANDED CODESHARE PARTNERS	16,731	238	14.23
5	**SOUTHWEST AIRLINES**	**11,443,404**	**53,660**	**4.69**
6	**DELTA AIR LINES NETWORK**	**10,049,935**	**52,388**	**5.21**
	- DELTA AIR LINES	7,728,303	41,390	5.36
	- BRANDED CODESHARE PARTNERS	2,321,632	10,998	4.74
7	**SPIRIT AIRLINES**	**1,212,719**	**6,708**	**5.53**
8	**JETBLUE AIRWAYS**	**1,237,415**	**7,871**	**6.36**
9	**UNITED AIRLINES NETWORK**	**7,771,841**	**58,469**	**7.52**
	- UNITED AIRLINES	4,823,435	35,542	7.37
	- BRANDED CODESHARE PARTNERS	2,948,406	22,927	7.78
10	**AMERICAN AIRLINES NETWORK**	**10,789,747**	**103,092**	**9.55**
	- AMERICAN AIRLINES	6,594,066	64,014	9.71
	- BRANDED CODESHARE PARTNERS	4,195,681	39,078	9.31
	TOTAL	47,961,192	303,410	6.33

* All U.S. airlines with at least 0.5 percent of total domestic scheduled-service passenger revenues.
(-) Data is not available for 2018. Comparison of 2020 and 2019 will appear in 2020 January-data ATCR.

Figure 4.4 Shows air travel consumer report: Mishandled baggage - Ranking of the U.S. reporting marketing carriers (monthly)

A study made by ANAC in 2015 showed that over 1 million domestic flights in Brazil, only 45% took off with average baggage more than 12kgs up to 23kgs (considered checked bag). The graph (Figure 4.4) below shows this proportion:

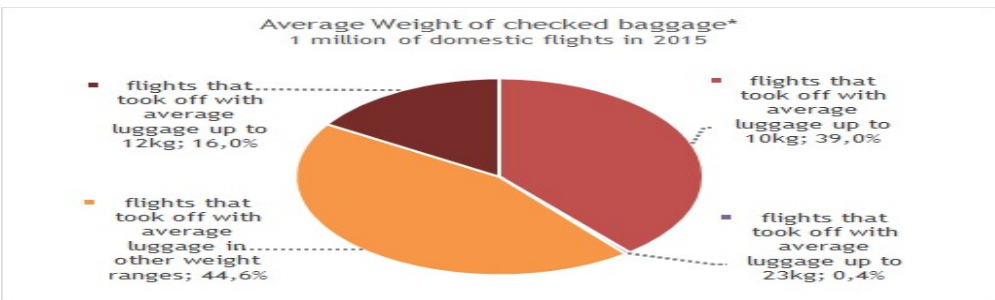


Figure 4.5 shows the average weight of checked baggage in 2015. Source: ANAC

Costa et al. Mishandled Bags Industry' Measure

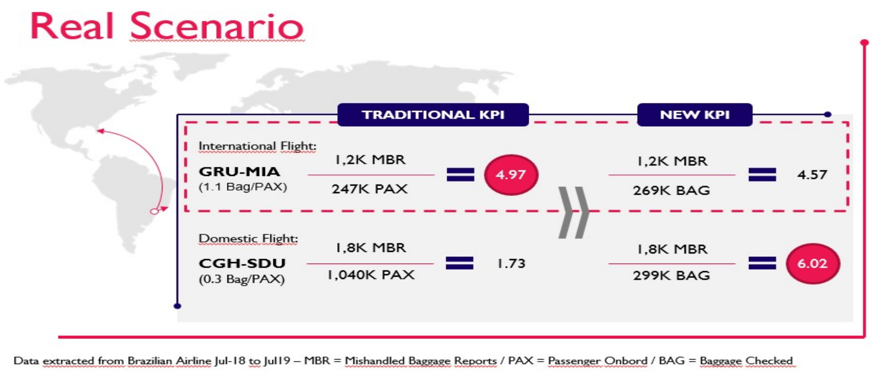
Adding the study, also according to ANAC, in 2016, only 30.8 million passengers had checked bags of a total of 88.1 million passengers on domestic flights in Brazil.

Another relevant point of this study, the number of checked baggage had decreased in the past few years; the image below (Figure 4.5) shows this tendency.

Figure 4.6 shows historical data of the average weight of baggage transported in Brazil in January. Source: ANAC

As seen above in the analyses, checked bags have a higher correlation than passengers. Therefore, using passengers as part of the metric, in this case, would end up in wrong assumptions and results.

Analyzing real data extracted from a Brazilian airline for different routes, we could see above the results:



We could see that calculated by the traditional key performance indicator; we had different results for CGH-SDU than in the new KPI; this happens mainly because it is a business route profile, which embraces our point that the customer behavior has changed and it is not possible making comparisons or track real numbers.

These numbers reinforce our purpose of changing and updating Brazilian airlines to the new KPI, as in the U.S. for the last few months.

Conclusions and Recommendations

This research motivation was based upon updating the KPI of mishandled baggage used worldwide. The current KPI measures passengers as a variable instead of checked bags, which is most correlated with the real root problem.
The most crucial change in the KPI is increasing customer experience and minimizing the risk of losing revenue.

Conclusions

The advantages of upgrading the KPI measurement of mishandled baggage from boarded passengers to checked baggage significantly outweigh the disadvantages. In addition to presenting a higher correlation in the performance of the simple linear regression analysis model, the new KPI will allow for improvements in baggage handling processes and will reduce costs with outsourced or own airline labor. The rapid growth in baggage tracking data technology through the adoption of RFID tagging, and the changing customer behavior that checks less baggage, further supports our thesis to update the mishandled KPI.

The researchers know that baggage handling is a sensitive item for customers, directly influences customer loyalty and experience, and impacts a brand's outcome and recommendations.

The effects of changing the KPI would be even more beneficial in a high-risk industry such as aviation. In Brazil, the companies face some particularities such as exchange rate volatility and excess regulatory standards that make the environment even more challenging.

Recommendations

Based on our findings and analyzing the big picture of world aviation, especially the Brazilian industry, our recommendation to airlines, airports, and industry partners is to transition from the current KPI model of mishandled baggage gradually. Despite the dramatic drop in mishandling rates in recent times, each mishandled bag is an additional cost for airports and airlines. There is still a lot of opportunities for more efficient ground operations and faster walking. Airlines must work directly with industry regulators, transportation departments, representative associations, and technology experts to put this issue on the agenda of industry discussions. According to our analysis and conclusions, the updating of this KPI would positively affect the results of the organizations. Two immediate benefits of this change are: adjustments of third-party contracts, making them more coherent regarding the fair amount of checked bags and improving the comparability of airlines and ground handlers' operations.

Another significant advantage in the remodeling of this KPI is higher accuracy and reality-related indicators; better will be the strategy to improve the service. So, consumers are likely to be the primary beneficiaries and can expect better service at a sensitive key point of the journey experience. In the final consequence, higher proportions of the happiest customers increase revenue for the industry and the companies.

Further researches should focus on the gradual increase in technology application in airport baggage handling operations. In addition to the industry-leading RFID technology, the trend in baggage handling is the gradual replacement for an automation-driven model. The research bet on better KPIs that translates into a long list of benefits, including reduced labor costs, increased accuracy, and increased customer service levels.

One of the greatest lessons learned is the continuing challenge of reconciling the speed of technological innovation with operational efficiency in reducing the number of lost baggage and consumer satisfaction. This involves, in addition to the financial issue for airlines and service providers, the real needs of customers and how much value the airline can add.

References

Aeronautics and Space. (2012, January 1). *Title 14 CFR II - Office of the secretary, department of transportation (Aviation Proceedings).* Retrieved from https://www.govinfo.gov/app/details/CFR-2012-title14-vol4/CFR-2012-title14-vol4-chapII

Air Travel Consumer Report Archive. (2012, August 28). Retrieved July 31, 2019, from https://www.transportation.gov/airconsumer/air-travel-consumer-report-archive

Airlines for America (2019, August 21). *A4A Presentation: Industry Review and Outlook.* Retrieved from http://airlines.org/dataset/a4a-presentation-industry-review-and-outlook/

Allon, G. (2009, December 10). *Ryanair: Fees for checked-in bags and … using the lavatories*.Retrieved from https://operationsroom.wordpress.com/2009/12/10/ryanair-fees-forchecked-in-bags-and-using-the-lavatories/#more-482

Amadeo, K. (2019, June 25). *How Congress Created the Greatest Bank Collapse Since the Depression*. Retrieved from https://www.thebalance.com/savings-and-loans-crisis-causes-cost-3306035

ANAC. (2015). *#aviaçãoparatodos*. Retrieved September 29, 2019, from http://www.transportes.gov.br/aviacaoparatodos/aviacao-para-todos

ANAC. (2019). *Boletim Trimestral de Monitoramento do consumidor.gov.br.* (1st QTR, pp. 01–14). Brasília, DF. https://www.anac.gov.br/consumidor/boletim-de-monitoramento. Retrieved from https://www.anac.gov.br/consumidor/boletins/1o-trimestre-2019

Arthur Zeikel. (1991). Forecasting and the Market. *Financial Analysts Journal*, 47(6), 15-18. Retrieved from http://www.jstor.org/stable/4479483

Aviation Strategy. (2018, December 14). *U.S. regionals: recession-proof airlines?* Retrieved December 23, 2019, from https://aviationstrategy.aero/newsletter/Dec-1998/3/US_regionals:_recession-proof_airlines?Issue number 14

Barlow, B. (2015, June 1). *The Ultimate Guide to Using Key Performance Indicators (KPIs) for Growth: Part 1*. Retrieved September 29, 2019, from https://beintheknow.co/key-performanceindicators-part1/

Biography.com. (2019, August 14). *Lyndon B. Johnson*. Retrieved from https://www.biography.com/uspresident/lyndon-b-johnson

Bowen, B. D., Headley, D. E., & Luedtke, J. R. (1992). A Quantitative Methodology for Measuring Airline Quality. *Journal of Aviation/Aerospace Education & Research*, 2(2). https://doi.org/10.15394/jaaer.1992.1063

Bureau of Transportation Statistics. (2017, February 14). *Airline Performance in 2016 - Some Highlights for Travelers*. Retrieved August 18, 2019, from https://www.transportation.gov/fastlane/airline-performance-2016-some-highlights-travelers

Bureau of Transportation Statistics. (n.d.). *Number 30 – Technical Directive: Mishandled Baggage, effective January 1, 2019*. Retrieved July 31, 2019, from https://www.bts.gov/topics/airlines-andairports/number-30-–-technical-directive-mishandled-baggage-effective-jan-1-2019

Carroll, L., & Gray, D. J. (1971). *Alice in wonderland*: Authoritative texts of alice's adventures in wonderland, through the looking-glass, [and] the hunting of the snark; backgrounds; essays in criticism. edited by donald J. gray (First ed.). New York: W. W. Norton.

Carter, S., & Cox, A. (2011, September 8). *One 9/11 Tally: $3.3 Trillion*. Retrieved from https://archive.nytimes.com/www.nytimes.com/interactive/2011/09/08/us/sept-11reckoning/cost-graphic.html?_r=1

Charnet, R. et al. (2018) *Análises de modelos de regressão linear: com aplicações*. 2 ed. Unicamp, Campinas, S.P.

Code of Federal Regulations. (2017, January 1). *Airline service quality performance reports*. Retrieved from https://www.govinfo.gov/app/details/CFR-2017-title14-vol4/CFR-2017-title14-vol4part234/summary

Creedy, S. (2019, April 25). *Airlines struggle to reduce global mishandled baggage rate*. Retrieved August 18, 2019, from https://www.airlineratings.com/news/airlines-struggle-reduce-globalmishandled-baggage-rate/

Deloitte University Press. (2017) *Rewriting the rules for the digital age*. Retrieved September 24, 2019, from https://www2.deloitte.com/us/en/insights/focus/human-capitaltrends/2017/introduction.html

Dempsey, P. (2003, March 23). *Aviation in Transition: Challenges & Opportunities of* ... Retrieved August 23, 2019, from https://www.icao.int/Meetings/ATConf5/Documents/dempsey.pdf

Drucker, P. (n.d.). *If you can't measure it, you can't manage it*. Retrieved August 18, 2019, from https://blog.marketculture.com/2009/03/20/if-you-cant-measure-it-you-cant-manage-it-peterdrucker/

Durcevic, S. (2019). *Business Intelligence: KPI Management And Best Practices: How To Find The Perfect KPI Solutions?*. Retrieved from https://www.datapine.com/blog/kpi-management-andbest-practices/

FAA (2017, January 4). *A Brief History of the FAA*. Retrieved from https://www.faa.gov/about/history/brief_history/#dereg

FORBES (2018, January 5). *How can businesses adapt to a rapidly changing world?*. Retrieved September 29, 2019, from https://www.forbes.com/sites/quora/2018/01/05/how-can-businessesadapt-to-a-rapidly-changing-world/#1a8a0e025930

House, C. (2006, May 3). *Mishandled Baggage: Problems and Solutions*. Retrieved October 21, 2019, from https://www.govinfo.gov/content/pkg/CHRG-109hhrg28283/html/CHRG-109hhrg28283.htm

IATA. (2013, October). *Ground Handling, International Air Transport Association*. Retrieved from https://web.archive.org/web/20131005235112/https://www.iata.org/whatwedo/opsinfra/Pages/ground-handling.aspx

IATA. (2018). *Global Passenger Survey*. Retrieved from https://www.iata.org/publications/store/Pages/global-passenger-survey.aspx

IATA. (2018). *IATA Business Case Report. End-to-End Baggage*. Retrieved from https://www.iata.org/whatwedo/ops-infra/baggage/Documents/end-to-end-baggage-businesscase-december-2018.pdf

IATA. (2018, December). *RFID: Back to the future*. Retrieved from https://www.airlines.iata.org/analysis/rfid-back-to-thefuture?_ga=2.25135738.1942166247.1567384144-138532754.1563761679

IATA. (2019). *Baggage Tracking*. Retrieved July 31, 2019, from https://www.iata.org/whatwedo/opsinfra/baggage/Pages/baggage-tracking.aspx

Kazmier, L. J. (1982) *Estatística aplicada à economia e a administração. 1* ed. São Paulo: McGraw-Hill do Brasil.

Leocha, C. L., & Travelers United. (2018, June 21). *Airline mishandled baggage statistics lie in favor of the airlines*. Retrieved July 31, 2019, from https://www.travelersunited.org/airline-mishandledbaggage-statistics-lie/

Maslow, H. A. (1962). *Toward a Psychology of Being*,Princeton,NJ

Maynard, M. (2008, May 21). *American Airlines to begin charging for checked luggage.* Retrieved August 22, 2019, from https://www.nytimes.com/2008/05/21/business/worldbusiness/21ihtamr.4.13100744.html

Mazareanu, E. (2019, April 29). *Causes of delayed bags in world aviation industry 2018*. Retrieved October 13, 2019, from https://www.statista.com/statistics/268277/causes-of-lost-luggage-in-airtravel/.

McDermott, J. (2017, June 26). *The History of Airline (De)Regulation In The United States*. Retrieved October 1, 2019, from https://aeronauticsonline.com/the-history-of-airline-deregulation-in-theunited-states/.

MELO, A. M. (2012) *Índices-padrão de indicadores econômico-financeiros das empresas de capital aberto do seguimento de construção civil integrantes do Novo Mercado*. Monografia (Especialização em Contabilidade e Controladoria Empresarial) – Universidade Estadual de Londrina, Londrina,PR.

Nicolae, M, Arıkan M, Deshpande V, Ferguson M, (2016). Do Bags Fly Free? *An Empirical Analysis of the Operational Implications of Airline Baggage Fees*. Retrieved from http://mentalfloss.com/article/85205/checked-baggage-fees-may-be-doing-you-favor Office of Airline Information, D. of T. B. of T. S. (2000, December 5). *General*. Retrieved from https://www.bts.gov/topics/airlines-and-airports/number-251-mishandled-baggage-reports Poleski, D. (2019, June 11). *Why are key performance indicators important?*. Retrieved September 20, 2019, from https://www.klipfolio.com/blog/kpi-importance

Power, Daniel J.. *Decision Support, Analytics, and Business Intelligence*, Second Edition, Business Expert Press, 2013. ProQuest Ebook Central, https://ebookcentral.proquest.com/lib/erau/detail.action?docID=1097922.

Reichheld, Frederick F. (2011). *The ultimate question 2.0: how net promoter companies thrive in a customer-driven world*. Boston, Mass.: Harvard Business Press.

Rhoades, D. L., & Waguespack, B., Jr. (2008). *Twenty years of service quality performance in the U.S. airline industry*. Managing Service Quality, 18(1), 20-33. doi:http://dx.doi.org.ezproxy.libproxy.db.erau.edu/10.1108/09604520810842821

Scotti, D., Dresner, M., & Martini, G. (2008). Baggage fees, operational performance and customer satisfaction in the U.S. air transport industry. *Journal of Air Transport Management,* 55(2016), 139–146. doi: https://doi.org/10.1016/j.jairtraman.2016.05.006.

SITA. (2018) Air Transport Industry Insights - *The Baggage Report*. Retrieved August 01, 2019, from https://dig.abclocal.go.com/wls/documents/2018/092118-wls-baggage-info-doc.pdf

SITA. (2018, May). *Baggage report. Ground Support Worldwide*, *26, 5*. Retrieved from http://ezproxy.libproxy.db.erau.edu/login?url=https://search-proquestcom.ezproxy.libproxy.db.erau.edu/docview/2042173210?accountid=27203

SITA. (2019). Baggage I.T. Insights report. *Do tracking points improve bag delivery?* Retrieved from https://www.sita.aero/resources/type/surveys-reports/baggage-it-insights-2019

SITA. (2019, August). *Baggage Report 2018*. Retrieved August 18, 2019, from https://www.sita.aero/resources/type/surveys-reports/baggage-report-2018

SITA. (2019). *Passenger I.T. Insights 20*19. Retrieved from https://www.sita.aero/resources/type/surveys-reports/passenger-it-insights-2019

SITA. (2019). *RFID for baggage tracking: SITA*. Retrieved from https://www.sita.aero/resources/type/white-papers/rfid-for-baggage-tracking

Statista, (Ed.). (2019, April). *Mishandled bags - cost to airline industry 2018*. Retrieved August 22, 2019, from htps://www.statista.com/statistics/268276/costs-for-the-airline-industry-due-tomishandled-bags-since-2007/

Stellin, S. (2013, July 8). *Making the airlines pay for lost luggage*. Retrieved from https://www.seattletimes.com/life/travel/making-the-airlines-pay-for-lost-luggage/

Transportation Department. (2011, July 15). *Reporting Ancillary Airline Passenger Revenues*. Retrieved from https://www.federalregister.gov/documents/2011/07/15/2011-17652/reporting-ancillaryairline-passenger-revenues

Transportation Department. (2016, November 02). *Reporting of Data for Mishandled Baggage and Wheelchairs and Scooters Transported in Aircraft Cargo Compartments*. https://www.federalregister.gov/documents/2016/11/02/2016-26181/reporting-ofdata-for-mishandled-baggage-and-wheelchairs-and-scooters-transported-in-aircraft-cargo

Unnikrishnan, M. (2015, June 4). *A Law That Changed The Airline Industry Beyond Recognition (1978)*. Retrieved from https://aviationweek.com/blog/law-changed-airline-industry-beyondrecognition-1978

Washington: Federal Information & News Dispatch, Inc. (2016). *Reporting of data for mishandled baggage and wheelchairs and scooters transported in aircraft cargo compartments*. Retrieved from ProQuest Central http://ezproxy.libproxy.db.erau.edu/login?url=https://search-proquestcom.ezproxy.libproxy.db.erau.edu/docview/1835008465?accountid=27203

Wayne, W. (2006). *Performance Dashboards: Measuring, Monitoring, and Managing your Business* (p.224), Hoboken, NJ: John Wiley and Sons publishers

DECISION SCIENCES INSTITUTE

Financial Evaluation for a Brazilian Airline to Conduct Components Repair in Brazil

Ana Moreno Sandoval
Ana.sandoval@latam.com
Embry Riddle Aeronautical University
Enzo Canavesi Luizetto
Enzo.luizetto@latam.com
Embry Riddle Aeronautical University
Joao Paulo Panssiera
Jao.panssiera@latam.com
Embry Riddle Aeronautical University
Klaus Goulart Brum
Klaus.brum@latam.com
Embry Riddle Aeronautical University
Dr. Leila Halawi
halawil@erau.edu
Embry Riddle Aeronautical University

ABSTRACT

The current economic world volatility and the increasing cost of fuel and security procedures established by the worldwide authorities force the air carriers to search for alternatives to reduce their operational costs. The maintenance costs are a significant subject for the air carriers. Therefore, any cost reduction obtained without maintenance service disruption may provide an opportunity to enhance the airline's competitiveness (Fritzsche, 2014). The purpose of this research is to evaluate and compare the technical and financial aspects of performing aircraft components repair in-house for a Brazilian airline versus outsourcing the servicing to a homologated company outside the country. In addition, this research will analyze the main challenges involved, such as costs involving the labor force, advantages and disadvantages of conducting the repair internally, the currency exchange and bureaucratic process applied by the Brazilian Customs to export and import components, and opportunities and companies classified as partners.

Keywords: Brazil airlines, repair, maintenance, brazil customs

INTRODUCTION

According to Vieira (2016), maintenance represents around 11% of the variable operating cost of any air carrier. In civil aviation, aircraft maintenance is a complex activity conducted by airline technical specialists, and this process usually occurs in the airport ramp areas, hangars, and shops. These designated areas provide the space and infrastructure needed to conduct the inspections, repairs, and overhaul, releasing the aircraft to operations. The direct maintenance costs refer to labor and materials spent on technical services performed by the mechanics in the aircraft and the components. Components maintenance costs in 2014 comprised 24% of the airline's direct maintenance costs per flight hour (IATA, 2015). Considering the period from 2010 up to 2014, the average cost per flight hour increased by 25% (IATA, 2015).

This research analyzes a Brazilian air carrier's technical and financial viability to expand its fundamental components repair capability using its Maintenance, Repair, and Overhaul (MRO) facility. The goal is to use its current structure and specialized maintenance personnel to develop new repair capabilities and compare the advantages and disadvantages of outsourcing to a repair station outside the country. Concerning the current procedure applied by the Air Carriers in Brazil, some of the fundamental aeronautical components are usually sent overseas to conduct such repairs. The process must strictly follow the Brazilian Customs export and import administrative requirements, including practical and active communication between importers and exporters and documentation filled out correctly. Any discrepancy identified by the customs authorities during the import or export process can cause a significant delay in the release of aircraft components and a substantial effect on the reputation and resulting profit margins. All evaluations of this research are based on a fictitious company named ABC Airlines that already provides components repair for the aircraft manufactured by Airbus. ABC Company Services does not perform a repair on engines nor auxiliary power units (APUs).

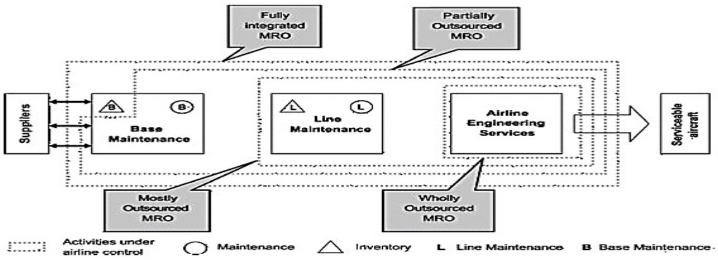
This research aims to identify opportunities to reduce costs to the Brazilian Air Carriers by conducting in-house repairs of aeronautical components instead of outsourcing the repairs outside the country. It is expected to result in significant savings to the companies.

Maintenance, Repair, and Overhaul (MRO)

The maintenance operation is a very dynamic task with activities segregated into planned, scheduled maintenance, a preventive service. On the other hand, unscheduled maintenance can be classified as a service not planned due to a component or an item failing or not working properly (Kinnisson, 2012). The maintenance and continuous standard operation processes are crucial to maintaining a high level of safety and reliability regarding the services involved (Kinnisson, 2012). According to Vieira (2016), maintenance, repair, and Overhaul – MRO involves retaining or restoring an item or a component, maintaining its standard performance following the requirements stated by the Aeronautical Authorities. This process involves a combination of technical characteristics and administrative management activities. If equipment or a component presents any mechanical discrepancies during the maintenance process, these parts should be removed from the aircraft and replaced with operative parts. Then, these parts were removed are sent to the repair premises (Yoon, 1994).

Figure 1 illustrates the complex maintenance process involving different MROs, considering four levels: Fully Integrated, Partially Outsourced, Mostly Outsourced, and Wholly Outsourced. Fully integrated MRO means the airlines perform all their maintenance activities internally (Al-Kaabi, 2007). Partially outsourced MRO means the airline conducts most of its maintenance needs internally, and only the base maintenance inventory is outsourced (Al-Kaabi, 2007). Mostly outsourced MRO means the airline managed its line maintenance and engineering services and outsource all base maintenance activities (Al-Kaabi, 2007). Wholly outsourced MRO means the airlines outsource all the MRO activities except the engineering services (Al-Kaabi, 2007).

Figure 1- MRO Model Depictions.



Commercial air carriers with MRO facilities use their maintenance capabilities to conduct fleet repairs and provide services for other airlines, consequently increasing their profit margin. It is common to see air carriers using their MROs as separate corporate units (Carpenter and Henderson, 2008). Due to the world aviation economic scenario complexity (such as currency exchange, political instability, etc.), airlines had to reconsider their maintenance, Repair, and overhaul-MRO strategies to provide specialized and diversified services. This will also enable the repair facilities to conform with their operational processes and meet their financial targets (Miroux, 2012).

Insourcing and outsourcing

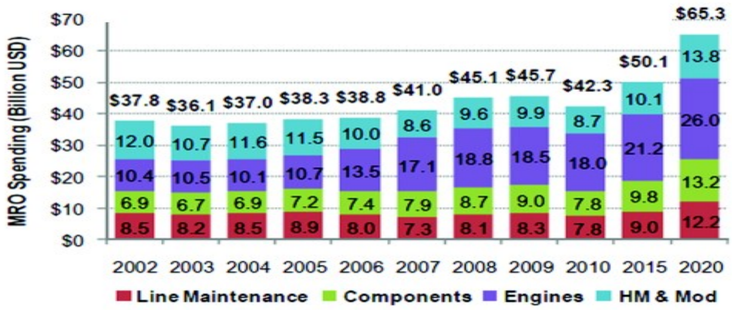
The reasons of whether or not to outsource repair components is an important and complex task and may influence decision-making (McIvor *et al.,* 1997); the use of outsourcing components maintenance allows the airline to expand its capabilities and resources without expanding its workforce (Lewis, 1999). In theory, some of the benefits of outsourcing are flexibility, pay per use concept, and low investment risks (Stapf, 2002). On the other hand, some disadvantages may appear, such as partnerships with wrong suppliers, difficulties monitoring contractor performance, or unexpected costs (Staff, 2002).

In contrast to outsourcing, the airlines with an MRO may use insourcing as a key factor of competitiveness to offer different in-house services to its fleet and third-party customers (Červinka, 2012). However, to maintain a sustainable and manageable insourcing option, essential subjects, such as costs, quality, lead times, inventory, and labor, must be aligned with the company's strategy and goals.

Maintenance Costs

According to Al-Kaabi (2007), maintenance in the aeronautical industry deserves special attention since it is a high-cost activity. The International Air transportation Association – IATA (2015) predicts the MROs spending to reach $ 65 billion in 2020, as shown in Figure 2

Figure 2- MRO Spending Forecast



Maintenance costs can range from 10% to 15% of the monthly operating costs (Samaranayake et al., 2002). These high costs are justified by the need for highly skilled labor and equipment involved, not to mention the high prices of the parts used in components or aircraft (Samaranayake et al., 2002). As obtained from the National Civil Aviation Agency (ANAC, 2018), Table 1 lists the operation costs of the five main Brazilian Air Carriers, considering the amount of R$ 34.5 billion spent in 2017.

Table1- Cost and Expenses of Brazilian Air Transportation

Description	%
Fuel and lubricants	27,5
Maintenance, rounding and insurances	20,3
People	17,4
Operating expenses	14,5
Miscellaneous costs	4,4
Navigation taxes	3,8
Ground handling	3,7
Depreciation	3,1
Passenger compensation	0,9
Judicial condemnation	0,9

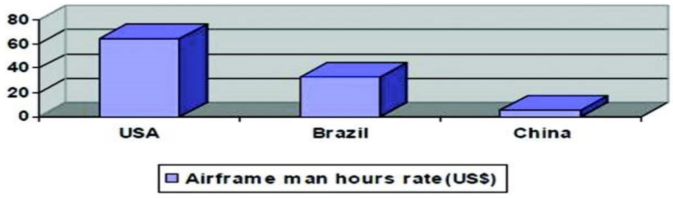
Although all procedures and techniques applied to maintenance hardly change, the maintenance costs are the opposite, and, consequently, it has been a significant issue for airlines. Thus, due to trimming costs and reducing investments, more airlines outsource MRO work rather than perform their aircraft maintenance (Christopher, 2007). Consequently, according to Christopher (2007), airline maintenance outsourcing has been growing, and the projection has an increased tendency, as demonstrated through Table 2.

Table 2 MRO Outsourcing

Year	%
1990	30
2000	50
2020	70

Considering labor costs, it is common that air carriers shift their heavy maintenance using overseas providers. Figure 3 shows the airframe person-hours rate comparing three countries. This information can assist in comprehending the impacts of labor on maintenance costs.

Figure 3- Estimated Man Hour Costs in Dollars



Supply Chain

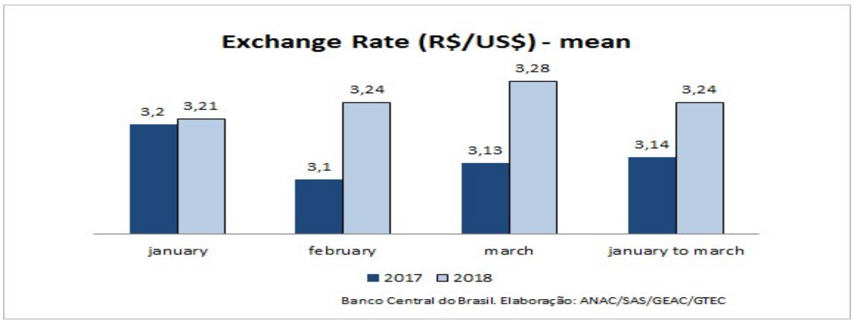
Supply Chain is considered one of the most complex sectors in the aeronautical industry, as Bales et al. (2004) mentioned. These complexities vary from certification requirements up to the handling and storage of the components. Besides the maintenance documentation, which is very important to assure the quality and safety of aircraft airworthiness, another relevant issue is the complex process of obtaining the qualification and authorizations for the suppliers to provide maintenance services (Vieira, 2016). Due to this complexity, there are few certified companies (Vieira, 2016). Some major suppliers decided to operate in different sub-tiers and became suppliers of their competitors (Vieira, 2016). This scenario may negatively affect the relationship between them and affect the end customers (Vieira, 2016). To guarantee safety and aircraft airworthiness, FAA, through Part 145, established all regulatory requirements for Certified/non-certified repair stations. **A vital** consideration punctuated by MacFadden (2012) was the total amount of money spending involved to keep inventory running, mainly regarding the intensive labor management involved. In addition, MacFadden said that this fact had been considered one vital matter to decide on outsourcing or in-house maintenance.

Exchange Variation

According to the Brazilian National Agency of Civil Aviation (ANAC) (June 29, the first quarter of 2018), the average exchange rate increased by 3.2% compared with 2017. The average price of aviation kerosene increased by 18.5%. The exchange rate exerts a strong influence on aircraft fuel, maintenance, and insurance costs, which accounted for 49.6%, half of the costs and expenses of public air services of Brazilian companies.

According to Figure 4, we can easily see the Real against the Dollar exchange rate increased. Thus, due to the strong influence of the dollar, on fuel, rental, maintenance, and aircraft insurance, these expenses represented 50% of the costs of air service in the first quarter of 2018.

Figure 4- Exchange Rate Affecting Airline Maintenance



Import Issues

The import defines any operation that facilitates the entry of goods in a customs territory after complying with the legal requirements and business (Assumpção, 2007). At the time of merchandise clearance, the import occurs in pre-established locations, with documents delivered at pre-established deadlines (Gama, 2013). In Brazil, the import documents have to be launched into the customs system called Siscomex. Once analyzed, Siscomex will identify the parameterization channel, where it will remain until the customs authority concludes, if applicable. There are four parameterization channels: green, yellow, red, and gray (Gama, 2013).

In the green channel, the Siscomex system will proceed to the automatic clearance of the merchandise (Assunpção, 2007). In the yellow channel, the documentation examination is carried out, and if nothing irregular is found, the process will be cleared without the need to examine the commodity (Assunpção, 2007). The merchandise will be removed in the red channel after the documentation is analyzed and its content is verified (Assunpção, 2007). In the gray channel, the clearance will only be made after the documentation examination, the verification of the merchandise, and the preliminary examination of the customs value (Assunpção, 2007). Brazilian economic and political instability generates impacts for air carriers due to frequent sector strikes. A governmental institution controls Brazilian customs, and it is being subjected to several strikes over the last years. The impact is a longer components turnaround time when repaired outside the country.

Airline Inventory Level

Airline punctuality is an essential requirement for customer satisfaction and brand reputation. Punctuality involves balancing several complex factors like the reliability of the fleet and components, Distribution and inventory level at the operational base,

crewmembers training standardizing the type of operation, and process effectiveness. Spare parts inventory exists to serve maintenance planned and unplanned activities. Based on an analysis conducted in 2012 by the International Air Transport Association (IATA) Maintenance Cost Task Force, the maintenance costs can be reduced by good planning. Excessive spare parts inventory leads to an immobilized asset, increasing costs, and impedes cash flows.
In contrast, inadequate spare parts can result in costly flight cancellations or delays, negatively impacting airline performance. In Brazil, it is critical to consider the increase in turnaround time for a component outsourcing repair outside the country due to the complexity, bureaucracy, and timing involved in the customs'. The timing expended in the customs process reduces the availability of serviceable spare parts in the company's inventory.

Most of the maintenance activities performed in an aircraft are planned. However, there are potential risks for some components to fail prematurely, and this results in additional impacts on operations. Airbus provided in March 2015 a mathematical model to calculate airlines' required inventory levels. This model uses a mathematical principle based on Poisson Distribution (Bethea, 1995). Poisson Distribution calculates the probability of an event to happen in a given time interval, and the result is independent of any other time interval. Airbus used Poisson distribution to calculate the stock level protecting the operation by making spare parts available and minimizing the investment by requesting a minimum resupply.

The operation of an airline has a demand profile for spare parts based on the equation:

$$D_{ann} = \frac{FH \; x \; FS \; x \; QPA}{MTBUR}$$

Where:

D_{ann}: Annual Demand

FH: Flight Hours per aircraft per year

FS: Number of Aircraft in the Initial Provision Period

QPA: Quantity per aircraft

MTBUR: Mean time between unscheduled removals for certain items A component resupply time can be calculated by the equation:

$$RST = \left[\left((MST + TT)x\left(1 - \frac{SCR}{1000}\right)\right) + ((LTM + AT)x\left(\frac{SCR}{1000}\right))\right]$$

Where:

RST: Re-Supply Time

MST: Max/Mean shop processing time

TT: Transit Time

SCR: Scrap Rate

LTM: Lead Time
AT: Administration Time

And, the inventory level can be calculated by the equation below:

$$D_{RST} = \frac{Dann}{365}\, x\, RST$$

Where:
D_{RST}: Expected Demand during Re-Supply Time

The final goal for an inventory analysis is to keep stock availability above the removal level, meaning parts are available when required. The Poisson distribution formula uses the probability of failure and recommends a quantity based on stock protection level.

The equation is:

$$P\{R \leq m\} = e^{-D_{RST}} x \sum_{0}^{m} \frac{(D_{RST})^m}{m!}$$

Where:

P: Probability R: Number of Removal m: Recommended Quantity

Poisson distribution calculation requires an iterative process assuming a recommended quantity based on stock level protection. Stock level protection is a risk assumed by the company that, following Airbus's recommendation, can vary from 80% to 98%, depending on the operational impact. The criticality of a component in operation is divided following the price of the component and the severity of the operational impact.

Reductions in the inventory levels and the amount of immobilized capital are strategic for an air carrier operation. Companies adopt strategies to maximize the return of capital applied. The goal is to obtain maximum return with minimum investment. From the operational perspective, this is equivalent to maximizing profit while minimizing assets. Therefore, an idea of demobilizing the assets gains strength within companies (Lima, 2003).

In the commercial aviation industry, the total value of spare parts inventories is around 45 million dollars, with an annual cost of opportunity estimated at 8 billion dollars (Sandvig and Allaire, 1998). This value is higher than the total profit obtained by the sector (Sandvig and Allaire, 1998). According to Oscar (2010), the aircraft components are repairable parts with high aggregated value, low annual removal rate, and low availability in the marketing, being difficult to be purchased. In Brazil, the procurement process is even more difficult, considering that aircraft parts are mostly imported, having few suppliers, and have a bureaucratic and complex customs process. Furthermore, the distance increases the storage and transportation costs.

Methodology

The development of an internal capability to perform maintenance on aeronautical components requires a technical and financial evaluation. The technical assessment considers the original equipment manufacturer's (OEM) component maintenance manual (CMM). It describes the technology involved, tools, equipment, bench tests, subparts, and staff training. All this information is part of the capability development for components repair. The required investment to buy tools, equipment, bench tests, and training are raised through quotations. The values provide the necessary information for net present value (NPV) calculation, which is part of the financial evaluation. The net present value (NPV) is the difference between the present value of cash inflows (internal repair) and the present value of cash outflows (external repair) over a period (Tse, 2017). The viability for an airline to internalize its component repair will depend on the forecast of the volume of components, the amount of investment, the complexity of the technology involved, and external component repair costs. A positive net present value indicates that the project earnings generated by the investment exceed the anticipated costs. In addition to the

NPV, the payback period is another important financial concept for investment return evaluation. The payback indicates the period (in years) to recover the initial investment (Tse, 2017). Another critical financial tool used to measure the project's viability is the Internal Rate of Return (IRR). This parameter is used to compare the discount rate of the invested capital with the cost of the capital. If the discounted rate is greater than the cost of capital, the investment is viable (Tse, 2017). IRR is the discounted rate when NPV is equal to zero.

Airbus recommends the use of Poisson Distribution to define inventory levels. This equation considers the risk assumed by the company to cover the operation. Normally, the risk taken can vary from 80% to 98%, dependent on the component operational impact (Airbus, 2015). The impact of a component to the aircraft operation is divided into three main categories per Airbus definition: go, go if, and no go.

This research considers a fictitious air carrier called ABC Airlines that operates in Brazil with a fleet of sixty aircraft manufactured by Airbus model A320. The sixty aircraft was chosen based on configuration and similarities involving the manufactures components by aircraft Air Transport Association (ATA-100) chapters. These similarities shall be considered to classify the sample size as valid for the research. The inventory level is calculated through Poisson Distribution and considers a centralized warehouse responsible for distributing spare parts to the other operational bases. The main goal of the mathematical model of this research is to estimate the inventory level reduction considering the turnaround time of components Repair house versus the turnaround time of outsourcing the component repair to a company outside the country.
Furthermore, the calculation considers the repair costs to perform maintenance in-house versus perform maintenance outsourcing. Christopher (2009) and Ballou (2006) explored the concept of a supply chain just in time when the inventory's size tends to zero. However, the customs bureaucracy, high costs of components, and the limited number of repair stations require a deep evaluation and investment in a safe inventory level to reduce operational impact.

This research considers a total population of 186 different components classified by the engineering team as low investment. Medium and high investments were not considered in this research. The focus is to reduce repair costs and inventory levels. All the 186 components are classified as "no go" or "go if," resulting in an operational impact on the company. ABC Airline evaluates how to reduce costs so that the research will analyze the costs for in-house repair versus the actual outsourcing repair costs.

Analysis of Inventory Level

The calculation of inventory level through Poisson Distribution considers the size of the fleet, annual removal rate, the number of installed components, reliability, and the desired level of operator protection. Airbus's recommended inventory protection level is 95% for the "go if" components and 98% for the "no go" components. The annual component removal rate and components reliability are based on historical information from ABC Airlines. The baseline is mainly focused on the information from the year 2017.

This research used Microsoft Excel to develop the Poisson Distribution Matrix to determine the amount of recommended inventory for all 65 components from the sample. The technical and financial evaluation considered five years where the components MTBUR is reduced 10% per year due to fleet aging. Furthermore, the research is based on a stable fleet of 60 aircraft and a component resale price at 60% of the full price of a new component. The price of the components is based on ABC Airlines. All calculations of Poisson Distribution are based on ABC Airline's historical components repair turnaround time that is 15 days' Repair plus three days logistics for the entire country. Therefore, the in-house repair turnaround time considered is 18 days. The outsourcing repair outside the country turnaround time was based on Airbus's recommendation for external repairs that is 60 days. These 60 days consider the exportation, repair, and importation process. According to ABC Airlines, this number can be highly increased from 60 to up to 120 days in case of a Brazilian customs strike, as observed in 2018. According to Airbus (2015), for parts where lead-time material is not available, the defaults to present values depending on the type of material should be considered, as mentioned in Figure 5

Figure 5- LTM default Table.

Type of material	LTM [days]
Airbus Proprietary Parts	10
Supplier Equipment	60
Supplier Equipment Breakdown Parts	60
Standard Hardware	30

Net Present Value and Payback

The internalization of components Repair for an airline depends on the technical expertise of the engineering team to develop the in-house capability and the correct determination of the investment amount in tools, equipment, bench tests, and training. The net present value and payback calculation considers all the investments and costs for an in-house repair and compares them with outsourcing the component repair outside the country. The total investment in tools, equipment, bench tests, and training is U$D 4,212,000.00. This value must be invested in the first year to guarantee the capability to perform the component repair. The initial investment in consumable materials for components repair is U$D 1,289,331.00. These materials are subparts as o-rings, packing, bolts, chemical materials, and several others to repair the components. The investment in consumable material is considered only in year zero because it represents an increase in its inventory. From years 1 to 5, the expenses on subparts are already considered the difference between the costs for in-house components repair versus outsourcing to a repair station outside Brazil. Historical data from ABC Airlines show that in-house repair is 35% cheaper when comparing to outsourcing. The main reason for this substantial difference between the in-house versus outsourcing repair costs is the cheapest workforce costs in Brazil compared to the US and Europe. Furthermore, the in-house repair does not consider the profit margin of the repair station because it is an internal gain.

The reduction in turnaround time considering the in-house repair (TAT 18 days) versus outsourcing repair (TAT 60 days) allows a decrease of approximately U$D 7,088,263.00 and a reduction in shipping handling expenses U$D 129,000.00. The total inventory reduction value was calculated considering the number of components available for sale after inventory level adjustment at 60% of the price of a new component. ABC Airline historical shows that 60% is a market price for used components to be sold as serviceable. Therefore, turnaround time is an essential factor in calculating inventory level through Poisson Distribution. Table 3 represents the five-year cash flow projections where investments are performed mainly in the first year and sales of overstock and due repair turnaround time reduction are divided into two years. Selling the overstock in two years will guarantee the operation and accommodate the in-house repair learning curve. Payback of the investment happens in the first year with an internal return rate of 54%.

Table 3- Five Years Projection Cash Flow

	ECONOMIC VIABILITY (x 1000)					
YEAR	0	1	2	3	4	5
CASH FLOW	$ -5.501,00	$ 4.596,00	$ 4.596,00	$ 1.052,00	$ 1.052,00	$ 1.052,00
	PAY IN					
IN-HOUSE GAIN		$ 923,00	$ 923,00	$ 923,00	$ 923,00	$ 923,00
SALE OF ASSETS		$ 3.544,00	$ 3.544,00	$0,00	$0,00	$0,00
HANDLING		$ 129,00	$ 129,00	$ 129,00	$ 129,00	$ 129,00
	PAYOUTS					
MATERIAL (INITIAL INVESTMENT)	$ 1.289,00	$ -	$ -	$ -	$ -	$ -
TOOLS	$ 4.212,00	$ -	$ -	$ -	$ -	$ -
DISCOUNTED CASH		$ - 4.178,00	$ - 3.798,00	$ - 790,00	$ - 718,00	$ - 653,00

Sandoval et al. **Evaluation Component Repairs**

FLOW		$ 2.889,00	$ 6.687,00	$ 7.478,00	$ 8.196,00	$ 8.850,00
PAYBACK	0,1					
COST OF CAPITAL (%)	$ 4.638,00					
NPV	54%					
IRR	1,84					
IP						

CONCLUSION

The total maintenance costs represent around 11% of the total air carrier operational costs. Maintenance is mainly divided into line, heavy, components, and engine maintenance. Components maintenance cost represents 24% of the total maintenance costs. An airline's operation depends on the availability of spare parts to support aircraft scheduled and unscheduled maintenance. The determination of inventory levels must balance the availability of spare parts to keep the operation and the amount of capital immobilized on inventory. Aeronautical components have a high aggregated value.

Airbus recommends using Poisson Distribution to calculate the inventory level to protect the airline operation due to an aircraft component's probability of failing. Poisson Distribution considers the turnaround time for a failed component to be repaired and returned to the company inventory. A longer turnaround time requires higher inventory to guarantee spare parts availability. Aeronautical components must be repaired on homologated repair stations. In Brazil, there are few repair stations with limited scope. Therefore, Brazilian Airlines are obligated to outsource several components maintenance to repair stations outside the country.

Outsourcing components maintenance to a repair station outside the country requires exportation and importation of the component through Brazilian Customs. This process is bureaucratic and requires several days to be cleared. Historical data shows an average of 60 days turnaround time for an outsourced component repair outside the country. The fictitious ABC Airline already has the capability to perform repairs in several components models. The main goal of this research was to evaluate the technical and financial elements involved in the expansion of the actual capability to repair components in-house. Historical data showed that in-house repair turnaround time is around 15 days.

The expansion of component repair capability involves a deep evaluation of the technical aspects that involve tools, equipment, bench tests, and training. The amount of investment required to buy tools, equipment, bench tests, and training was raised through market quotations already available at ABC Airlines. The financial evaluation considered the amount of capital invested in tools, equipment, bench tests, and training, the projected in-house repair costs, the volume of repairs per year, and the possibility of inventory reduction due to turnaround time reduction versus the outsource repair costs. This research provided elements on how to calculate the reduction in inventory levels, keeping the operation protected. Furthermore, it showed the mathematical model used to measure the financial viability to perform aeronautical components in-house repair for a Brazilian airline.

RECOMMENDATIONS

The development of components repair capability is based on the Original Equipment Manufacturer (OEM) Component Maintenance Manual (CMM). This manual brings all the technical information required, as tools, equipment, bench tests, consumable materials, and others. There is significant commercial interest from the OEM to keep control of the technology and sell the repair services by themselves. In several cases, the price of the tools, equipment, and bench tests to develop in-house capability result in an unviable financial analysis.

The investment related to capability development can be considerably reduced when the repair station has a good and specialized development engineering team. This team is responsible for evaluating the required tools, equipment, and bench tests and developing these devices internally. The internalization of components repair requires good consumable materials planning. Every aeronautical component is composed of several subparts that must be available in stock for replacement. Subparts prices are low when compared to the price of the component. However, the lack of some subparts will not allow the full assembly of the component, stopping the final release of the aeronautical part.

The technology involved in components repair and its intrinsic complexity requires constant and recurring technician training. The learning curve for a mechanic takes around two years to get the necessary experience. The mechanic's expertise keeps increasing over the years. Therefore, the company must be cautious with employees' turnover and create means to keep the engagement. The aeronautical authority monitors airlines' On-Time Performance in Brazil, and it has a strong relation with reputation. Aircraft's reliability has a close relation to components reliability.

The internalization of components repair is an important factor to monitor and increase components reliability. The engineering team can evaluate components' degradation and failure modes and propose product and process improvements. For future research, we could consider using Poisson for spare parts to reduce the number of parts in the inventory.

Another point to be considered a potential to the aviation industry is strengthening airline maintenance internal capability through customized benches, tools, and reverse engineering. To maximize the benefits of this research and maintain insourcing maintenance in Brazil, the approval of repairs beyond those authorized and stated in the manufacturers' manuals should be considered. The manufacturers restrict some repairs to protect themselves, forcing the airlines to send their components to them. This approval might reduce the level of airline inventories.

For further research, the airlines can conduct a benchmarking with other companies, considering the same peculiarities and operational complexity as Brazil, mainly regarding the customs process and the impact of exchange variation and evaluating the feasibility to provide repair services. The financial evaluation was performed considering the actual Brazilian political and economic scenario. A sensitivity analysis was not performed considering projections. This analysis can be part of further research considering the influence of rate exchange variation, repair turnaround time impact, and component reliability. Future studies can evaluate how much these parameters have an impact on in-house repair financial viability.

In the end, this research did not consider cost opportunities for payback calculation. Consequently, it represents an opportunity to be better explored in further researches. When necessary to keep the aircraft operating, the company warehouse's availability influences financial evaluation. The financial analysis of this research considered only the inventory reduction impact; however, the internalization has a bigger financial gain when considering the operational impact due to flight cancellations and flight delays. This research doesn't cover this parameter generating an opportunity for further studies in this subject.

REFERENCES

Al-kaabi, H., Potter, A., & Naim, M. (2007). An outsourcing decision model for airlines' MRO activities. Journal of Quality in Maintenance Engineering, 13(3), 217-227. doi:10.1108/13552510710780258

Airbus Service Solutions (2015). "Mathematical Model."

Airbus Maintenance Manual (2018). "Airbus Flight Operational Manual."

Assunpção, R.R. 2007, ¨Exportação e Importação – Conceitos e Procedimentos Básicos¨. 1. Ed. São Paulo.

Bales, R. R., Maull, R. S. and Radnor, Z. 2004, ¨The development of supply chain management within aerospace manufacturing sector¨. Supply Chain Management, 9 (3), 250-255.

Bazargan, M., et al. (2013). "A Simulation Approach to Airline Cost-Benefit Analysis¨.

Bazargan, M. (2016). "Airline Maintenance Strategies – in-House vs. Outsourced – an Optimization Approach." Journal of Quality in Maintenance Engineering, 22 (2), 114–129., doi:10.1108/jqme-08-2015-0038

Bethea, R., et al. "Statistical Methods for Engineers and Scientists ¨.

Carpenter, R., & Hendeson, A. ¨Keep Them Flying: Find your winning position in the MRO game. IBM Global Business Services: https://www-935.ibm.com/services/us/gbs/bus/pdf/gbw03020-usen-00_keep_flying.pdf

Cochran, W. G. (1977). Sampling techniques (Third ed.). New York: Wiley.

Fritzsche, R., Gupta, J. N. D., & Lasch, R. (2014). Optimal prognostic distance to minimize total maintenance cost: The case of the airline industry. International Journal of Production Economics, 151, 76-88. doi:10.1016/j.ijpe.2014.02.001

Hassanain, M. A., Assaf, S., Al-Hammad, A., & Al-Nehmi, A. (2015). A multi-criteria decision-making model for outsourcing maintenance services. Facilities, 33(3/4), 229-244. doi:10.1108/F-01-2013-0003

IATA (2015). Best Practices for Component Maintenance Cost Management from https://www.iata.org/whatwedo/workgroups/Documents/MCTF/CMCM-2nd-edition.pdf

International Trade Administration (2016). Top Markets Report Aircraft Parts Country Case Study Brazil from https://www.trade.gov/topmarkets/pdf/Aircraft_Parts_Brazil.pdf

Kinninson, H. A. 2012. ¨Aviation Maintenance Management¨. McGraw-Hill, New York, NY.

Lewis, B.T. (1999), Facility Manager's Operation and Maintenance Handbook, McGraw-Hill, New York, NY.

McFadden, M., & Worrells, D. S. (2012). Global outsourcing of aircraft maintenance. Journal of Aviation Technology and Engineering, 1(2), 4.

McIvor, R. T., Humphreys, P. K., & McAleer, W. E. (1997). A strategic model for the formulation of an effective make or buy decision. Management Decision, 35(2), 169-178. doi:10.1108/00251749710160331

Miroux, F, Y. J. 2012. ¨How to get sustainable maintenance spare parts management?¨ Thesis, Master of Science in Operations Management and Logistics, TU/e

Paul Yoon, K., & Naadimuthu, G. (1994). A make-or-buy decision analysis involving imprecise data. International Journal of Operations & Production Management, 14(2), 62-69. doi:10.1108/01443579410053239

Samaranayake, P., Lewis, G. S., Woxvold, E. R. A., & Toncich, D. (2002). Development of engineering structures for scheduling and control of aircraft maintenance. International Journal of Operations & Production Management, 22(8), 843-867. doi:10.1108/01443570210436172

Stapf, J. (2002). The advantages and disadvantages of outsourcing. Journal - American Water Works Association, 94(4), 66-68. doi:10.1002/j.1551-8833.2002.tb09454.x

Vasigh, B., & Gorjidooz, J. (2017;2016;). Engineering economics for aviation and aerospace. Abingdon, Oxon; New York, NY;: Routledge is an imprint of the Taylor & Francis Group, an Informa Business. doi:10.4324/9781315644257

Vieira, D. R. (2016, February 12). Maintenance, Repair, and Overhaul (MRO) Fundamentals and Strategies: An Aeronautical Industry Overview. Retrieved July

14, 2018, from https://www.ijcaonline.org/research/volume135/number12/vieira2016-ijca-908563.pdf

Tse, T. C. M. (2017). Corporate finance: The basics (1st ed.). Milton: Routledge Ltd. doi:10.4324/9781315526577

Yamane, Taro. 1967. Statistics, An Introductory Analysis, 2nd Ed., New York: Harper and Row

DECISION SCIENCES INSTITUTE

Student Masters Projects using Excel Routing Templates with AllDifferent Constraint

Rick Hesse, Professor Emeritus, Pepperdine University, Adjunct Professor, Union College (KY)
rickhesse@aol.com

ABSTRACT

Three different Traveling Sales Routing Excel templates are shown with examples of several working professional master's students' term-end projects in real business situations. Optimal solutions require n^2 variables, which limits problems of 14 nodes or less. Algorithms with the AllDifferent constraint in Excel's Solver only need *n* variables (up to 200 variables) with the caveat that the solution may not be optimal and there is no way of determining if the solution is close to optimal. For most real problems, if the quick & dirty solution is better than anything found so far, it is good enough.

KEYWORDS: Routing, Student, Project, Heuristic, AllDifferent

INTRODUCTION

My Management Science classes for graduate students makes heavy use my own Excel templates, and at the end of the semester each student must choose a real or realistic problem to demonstrate their efficacy in their workplace. Since these students were gainfully employed, I have seen many practical applications of these templates and learned more about better ways to gain a good solution using the AllDifferent constraint in Solver for routing problems. Three Traveling Sales Routing Excel templates are shown with examples of working professionals using the Solver taught in my general MBA (Pepperdine) and Data Analytics MSBA (Lincoln Memorial University) classes over the years. I require students at the end of the semester to use 2 or more templates for individual working graduate student term-end projects in their real business situations. A popular set of templates with students are using those which involve routing problems. These three templates are Traveling Sales Tours (**TST**) which are a complete loop, Traveling Sales String (**TSS**), which does not close the loop but visits each location, and Delivery Problems (**DP#n**) where n is 2 or more loops starting and ending from the home base. Using optimal assignment algorithms for each of these requires at least n^2 variables, which limits the regular Solver to 14 nodes or $14^2=196$ variables since only 200 variables are allowed. By employing the AllDifferent constraint in my templates, only n variables are required but the solution is not guaranteed to be optimal. However it takes only a few minutes for each trial solution and a good solution can be found that satisfies each situation.

THE AllDifferent CONSTRAINT

"The AllDifferent *constraint is a specialized constraint which forces every decision variable in a given group to assume a value different from the value of every other decision variable in that group. ... In other words, no two of those decision variables will have the same integer value when this constraint is satisfied."*

Adding "dif" for the type of constraint for one or more variables in Excel, starts a heuristic branch and bound technique named "AllDifferent" to try different combinations of 1-2-3-->n for the unique values of selected variables. For a moderate-sized problem, a standard laptop computer can generate 20,000 possibilities in just a few seconds. This can be very handy for routing problems, which can have millions or more possible routes. Make sure that you also choose the Solver to run **Evolutionary Solver**.

TRAVELING SALES TOURS

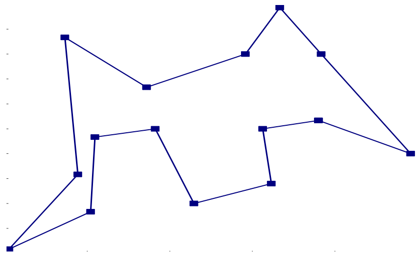


Figure 1. *Traveling Sales Tour*

The graph in **Figure 1** shows 15 nodes (cities, stops, locations) being visited once and only once and returning to the first node. It makes no difference where this starts, as it is a complete loop, and you can then reorder the sequence as desired. The object is to minimize distance, cost, time, or other measures and find the minimum tour. If the distances between nodes is not symmetric, there are (n-1)! or in this case, 14! possibilities, or 87,000,000,000 or 8.7E+10 possible routes. If 10,000 could be computed per second, it would take about 2,764 years to find the best of all those routes. Using the assignment algorithm and eliminating subtours (Hesse, R, 2021, Chapter 4) would require 152 variables or 225, which exceeds the capacity of the standard Solver in Excel. Thus the need for the AllDifferent constraint.

Student Project #1: Numero Uno Pizza (Teemu)

Teemu Kivviko was an MBA student and an accountant for Numero Uno Pizza, which had 24 locations besides the Office, and he needed to drop off pay checks (took about 10 minute each stop) once a month on payday. With so many part-time staff and last-minute changes, these bundles of checks needed to be hand delivered. Shown in **Figure 2** is the setup for the problem. This file is included with conference materials for your examination.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1	NUMERO UNO PIZZA VISITS												MPH		Dropoff		min/hr														
2	NUP405b												35		10		60														
3	#	Store	Store	Miles		Miles	Office	3	5	6	8	9	10	11	16	17	21	22	26	27	28	29	35	43	47	57	59	71	72	74	401
4	1	Office	3	16		Office	999	16	16	9	8	21	51	18	19	9	22	37	23	27	2	28	21	23	12	7	4	7	1	31	41
5	2	3	5	6		3	16	999	6	15	9	29	63	11	26	15	7	24	8	17	17	41	27	7	3	19	20	16	16	21	56
6	3	5	6	13		5	16	6	999	13	10	29	64	13	27	15	8	27	8	12	17	41	25	10	11	17	20	13	16	24	54
7	4	6	8	18		6	9	15	13	999	18	29	56	22	26	21	19	36	19	23	7	33	15	19	16	4	10	2	9	33	44
8	5	8	9	20		8	8	9	10	18	999	20	57	13	17	6	18	33	18	22	9	34	29	18	8	18	13	11	8	29	47
9	6	9	10	54		9	21	29	29	29	20	999	54	27	4	15	36	50	36	40	26	30	37	35	26	27	21	33	21	45	44
10	7	10	11	64		10	51	63	64	56	57	54	999	64	57	54	70	85	70	75	53	24	44	71	60	53	47	59	51	81	21
11	8	11	16	24		11	18	11	13	22	13	27	64	999	24	14	12	23	16	25	19	42	35	8	7	27	22	23	18	19	55
12	9	16	17	12		16	19	26	27	26	17	4	57	24	999	12	32	46	33	37	22	34	34	32	22	23	17	30	19	41	47
13	10	17	21	21		17	9	15	15	21	6	15	54	14	12	999	21	34	22	26	13	31	29	20	11	18	12	24	9	30	45
14	11	21	22	17		21	22	7	8	19	18	36	70	12	32	21	999	17	4	13	24	47	32	3	15	23	27	20	22	22	61
15	12	22	26	16		22	37	24	27	36	33	50	85	23	46	34	17	999	16	20	38	62	48	17	25	40	41	36	37	27	77
16	13	26	27	7		26	23	8	8	19	18	36	70	16	33	22	4	16	999	7	24	48	32	7	11	24	27	20	23	26	61
17	14	27	28	28		27	27	17	12	23	22	40	75	25	37	26	13	20	7	999	28	52	36	22	22	28	32	24	27	36	65
18	15	28	29	29		28	2	17	17	7	9	26	53	19	22	13	24	38	24	28	999	29	20	24	14	4	6	5	2	34	42
19	16	29	35	23		29	28	41	41	33	34	30	24	42	34	31	47	62	48	52	29	999	23	48	37	30	24	37	28	58	12
20	17	35	43	32		35	21	27	25	15	29	37	44	35	34	29	32	48	32	36	20	23	999	32	29	12	17	17	21	46	30
21	18	43	47	12		43	23	7	10	19	18	35	71	8	32	20	3	17	7	22	24	48	32	999	12	23	27	20	23	18	61
22	19	47	57	22		47	12	3	11	16	8	26	60	7	22	11	15	25	11	22	14	37	29	12	999	22	16	20	12	23	51
23	20	57	59	7		57	7	19	17	4	18	27	53	27	23	18	23	40	24	28	4	30	12	23	22	999	7	8	7	37	40
24	21	59	71	14		59	4	20	20	10	13	21	47	22	17	12	27	41	27	32	6	24	17	27	16	7	999	14	4	38	38
25	22	71	72	7		71	7	16	13	2	11	33	59	23	30	24	20	36	20	24	5	37	17	20	20	8	14	999	7	33	46
26	23	72	74	31		72	1	16	16	9	8	21	51	18	19	9	22	37	23	27	2	28	21	23	12	7	4	7	999	31	41
27	24	74	401	72		74	31	21	24	33	29	45	81	19	41	30	22	27	26	36	34	58	46	18	23	37	38	33	31	999	72
28	25	401	Office	41		401	41	56	54	44	47	44	21	55	47	45	61	77	61	65	42	12	30	61	51	40	38	46	41	72	999
29		Total TSP Miles		606																											

Figure 2. *Numero Uno TST Setup*

Cells **A4:A28** contain the starting values of the AllDifferent constraint, shown in the setup in **Figure 3**. Note that the constraint identifies the Changing Variables as "**dif**" which changes to

AllDifferent in the Solver setup once selected. Note also that the type of Solving Method selected is Evolutionary.

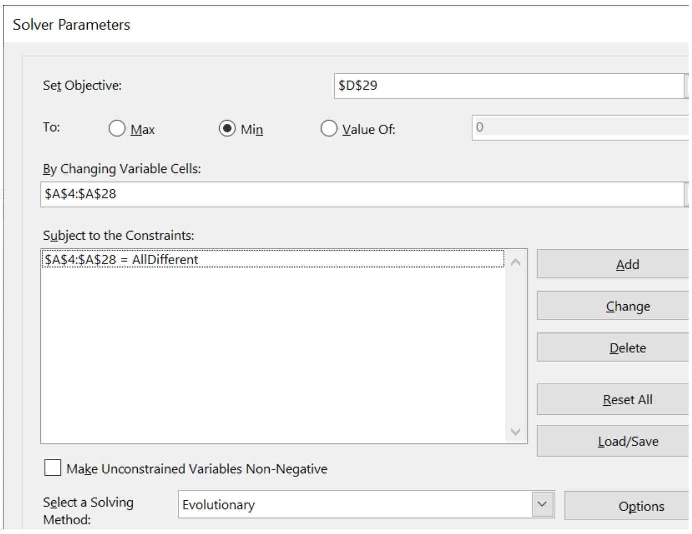


Figure 3. *Solver Setup*

The formulas of interest for the columns B, C and D are:

B4: =**@INDEX**($F$4:$F$28,A4)
C4: =B5
D4: =**@INDEX**($G$4:$AE$28,A4,A5)

And the last row (28)
D28: =**@INDEX**($G$4:$AE$28,A28,A4)
points back to the beginning.

The solution is shown in **Figure 4** and then the Report next to it shows the distance, time and complete TST. This is not guaranteed to be optimal, and it makes sense to make a copy of the setup and enter different orders to numbers 1➔25 (backward, every other one, etc.) to see if a better tour exists. This solution in the Report shows Teemu leaving the Office to go to store #59, then #72, etc. and ending up at the office with no need for 10 minutes there. Done in less than 12 hours when all stores are open.

#	Store	Store	Miles
18	43	21	3
11	21	22	17
12	22	27	20
14	27	26	7
13	26	5	8
3	5	3	6
2	3	47	3
19	47	8	8
5	8	Office	8
1	Office	59	4
21	59	72	4
23	72	28	2
15	28	71	5
22	71	6	2
4	6	57	4
20	57	35	12
17	35	401	30
25	401	10	21
7	10	29	24
16	29	9	30
6	9	16	4
9	16	17	12
10	17	11	14
8	11	74	19
24	74	43	18
	Total TSP Miles		285

REPORT

From Store	To Store	Miles	Drive Min	Dropoff Min		
Office	59	4	6.9	10		
59	72	4	6.9	10		
72	28	2	3.4	10		
28	71	5	8.6	10		
71	6	2	3.4	10		
6	57	4	6.9	10		
57	35	12	20.6	10		
35	401	30	51.4	10		
401	10	21	36.0	10		
10	29	24	41.1	10		
29	9	30	51.4	10		
9	16	4	6.9	10		
16	17	12	20.6	10		
17	11	14	24.0	10		
11	74	19	32.6	10		
74	43	18	30.9	10		
43	21	3	5.1	10		
21	22	17	29.1	10		
22	27	20	34.3	10		
27	26	7	12.0	10		
26	5	8	13.7	10		
5	3	6	10.3	10		
3	47	3	5.1	10		
47	8	8	13.7	10	Total	
8	Office				Time	
Total TST Miles		277	474.9	240.0	715	11:54

Figure 4. *Best Solution and Time Report*

Teemu saw that this solution allowed him to deliver all the paychecks within the 12-hour window of all the stores being opened and was a possible solution.

TRAVELING SALES STRING (TSS)

A Traveling Sales String **(TSS)** shown in **Figure 5** does not connect back to the start but does visit each city (or stop) once and only once. Two ways will be shown to determine the minimal length, time or cost. The dotted arrow represents not including the value of the last link.

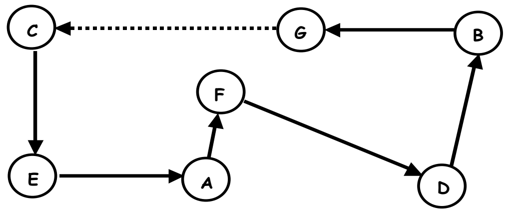


***Figure 5**. Traveling Sales String*

Student Project TSS (Rick)

As a graduate student (yes, I used to be a student), my Master's Thesis was solving a problem for Monsanto Chemical Company about seven (7) batches of different chemicals to be produced in the same large vat. Thus rather than distances, it was the change-over cost of cleaning the vat from using chemical 1 to chemical 2 that was involved. The thesis was much more involved because sometimes two or more batches of the same chemical needed to be produced but not in succession. This example for my students made a good homework problem for **TSS**, both to solve adding constraints to the assignment problem to eliminate loops and as a heuristic with the AllDifferent constraint. The cost of cleaning up from chemical 1 to produce chemical 2 is shown in **Figure 6**. The optimal **TST** is found and then the highest cost link can be eliminated to get a good **TSS**. The cost on the diagonal is set much higher than all other costs to keep the Solver from choosing not to leave 1-1 or 2-2, etc.

Using the Assignment algorithm has its limitations for this type of problem because at most 14 chemical (stops, cities, etc.) can be used which would produce 14X14 or 196 variables and there is a limit of 200 or less for variables. The Premium Solver would have to be employed for larger problems.

CHANGE-OVER COSTS

CV405ab

CHANGE-OVER COST BETWEEN BATCHES

	COST	1	2	3	4	5	6	7
	1	$9,999	$400	$800	$600	$500	$800	$700
F	2	$700	$9,999	$900	$1,000	$800	$200	$1,400
R	3	$500	$300	$9,999	$1,200	$900	$1,100	$700
O	4	$600	$200	$900	$9,999	$800	$700	$500
M	5	$800	$600	$500	$800	$9,999	$1,000	$1,100
	6	$200	$400	$900	$500	$1,000	$9,999	$300
	7	$900	$1,000	$800	$900	$1,100	$500	$9,999

ALLOCATION TO JOB

	JOB	1	2	3	4	5	6	7	
	1					1			1
F	2						1		1
R	3							1	1
O	4		1						1
M	5			1					1
	6	1							1
	7				1				1
		1	1	1	1	1	1	1	$3,200

<= Optimize

Break Tour ≤ k-1			Cost
2	2	1-5-3-1	$3,100
1	1	4-7-4	$3,100

Assignment Report

From	To	Cost
1	5	$500
2	6	$200
3	7	$700
4	2	$200
5	3	$500
6	1	$200
7	4	$900
	TOTAL	$3,200

Figure 6. *Optimal TST Solution w/Assignment*

The optimal **TST** is $3,200 but it does not need to clean up from the last chemical to start over because basic maintenance will take care of that. So eliminating the most costly branch of the TSS (7➔4) saves $900, shown as the first two parts of **Figure 7**.

Optimal traveling salesman route

From	To	Cost
1	5	$500
5	3	$500
3	7	$700
7	4	$900
4	2	$200
2	6	$200
6	1	$200
	TOTAL	$3,200

Quick & Dirty String using Optimal TST

From	To	Hours
4	2	$200
2	6	$200
6	1	$200
1	5	$500
5	3	$500
3	7	$700
	TOTAL	$2,300

Quick & Dirty String using AllDifferent

From	To	Cost
5	3	$500
3	2	$300
2	6	$200
6	1	$200
1	4	$600
4	7	$500
	Total Cost	$2,300

Figure 7. *TSS for Monsanto Problem*

	A	B	C	D	E	F	G	H
4	#	Job	Job	Cost		COST	1	2
5	5	5	3	$500		1	$9,999	$400
6	3	3	2	$300		2	$700	$9,999
7	2	2	6	$200		3	$500	$300
8	6	6	1	$200		4	$600	$200
9	1	1	4	$600		5	$800	$600
10	4	4	7	$500		6	$200	$400
11	7	7				7	$900	$1,000
12		Total Hours		$2,300				

Figure 8. *AllDifferent TSS*

The third part of **Figure 7** shows the use of the heuristic **TSS** with the AllDifferent constraint and gives an alternate solution, which is also optimal, but if we did not solve the problem first with the Assignment algorithm, we would not know that it was optimal. **Figure 8** shows the AllDifferent heuristic solution template result.

The algorithm for the TSS shown in **Figure 8** is almost identical to the one for TST, but the last two cells in **C11:D11** are blank and do not connect back to the beginning of the tour. This result was achieved by entering 1-2-3-4-5-6-7 in **A5:A11**. Run the Solver as long as you want and you can also try any other numbers in **A5:A11** and see if a better solution is found.

Professor Project (Rick): Reduce the number of two-in-a-row exams

When the author was a civilian professor at the U S Military Academy (West Point), he developed an algorithm to schedule the cadet term-end (final) exam schedule. There were 11 exams periods (3-3-3-2) scheduled Monday through Thursday. Cadets could not take two exams at the same time so first an algorithm was developed to reduce the number of conflicts to a manageable number and officers from the dean's office would work these out. (Hesse, Weinstock, Grum, 1982).

But if a cadet was scheduled for three exam periods in a row, that would also be resolved, with make-ahead or make-up exams. This last condition was partially ameliorated by reducing the number of 2 exams in a row in the hopes that doing so would drastically cut down the probability of 3 exams in a row. **Figure 9** shows the matrix of the number of students with back-to-back (B2B's) exams for all 11 periods. The **TST** solution with the largest branch has 9 B2B's on the left of **Figure 9**. The resulting TSS report is on the right of **Figure 8** and shows just 27 B2B's and hopefully only 2 or 3 of those would be three-in-a-row.

	From Period	To Period	B2Bs
4	4	9	2
9	9	1	5
1	1	5	2
5	5	6	0
6	6	2	0
2	2	3	2
3	3	10	2
10	10	8	4
8	8	11	9
11	11	7	7
7	7	4	3
4			36

BACK-TO-BACK (B2B)

	Perioc	1	2	3	4	5	6	7	8	9	10	11	0
	1	99	6	22	8	2	13	8	15	5	11	17	9
F	2	6	99	2	7	8		6	12	4	8	16	0
	3	22	2	99	6	3		4	6	8	2	6	5
	4	8	7	6	99	8	7	3	11	2	7	11	5
R	5	2	8	3	8	99		10	9	6	7	14	3
	6	13			7		99	4	8		5	3	8
	7	8	6	4	3	10	4	99	9	4	6	7	22
O	8	15	12	6	11	9	8	9	99	10	4	9	7
	9	5	4	8	2	6		4	10	99	6	11	4
	10	11	8	2	7	7	5	6	4	6	99	8	3
M	11	17	16	6	11	14	3	7	9	11	8	99	12

Traveling Sales String

From Period	To Period	B2B's
11	7	7
7	4	3
4	9	2
9	1	5
1	5	2
5	6	0
6	2	0
2	3	2
3	10	2
10	8	4
		27

Figure 9. *TST B2B Solution w/TSS Report*

Both heuristics found that 27 B2B's were the fewest that could be found and the same sequencing. interestingly, there can two different orderings, backwards and forwards. This is not surprising when you realize that the matrix of B2B's must be symmetrical, because the B2B's for two periods (i,j) is the same as (j,i).

DELIVERY PROBLEMS, DP#2, (Reda)

Reda Alhoussayni, an MBA candidate (and also a pharmacist), had an opportunity to become a Pharmacy District Manager with Ralphs grocery stores. Management wanted to build a pharmacy in every store and the first area of interest was in northern California, shown in **Figure 10**. Reda wanted to get an idea of how much driving he would have to do to audit these stores. He currently was employed in southern California and working on his MBA at night. He obtained information about the eleven stores he would be working with and using MapQuest got a good idea of the driving distances between stores and set up the following template, shown in **FIGURE 11**. The AllDifferent constraint can be employed along with the Evolutionary Solver as a heuristic for the Delivery Problem. To accomplish this, we must augment the table of times by adding a duplicate row and column for each Base needed (in this case, two of them) and name them Oakland1 and Oakland2. The values of 999 keep the assignment problem from choosing to stay in a city rather than force it to move on.

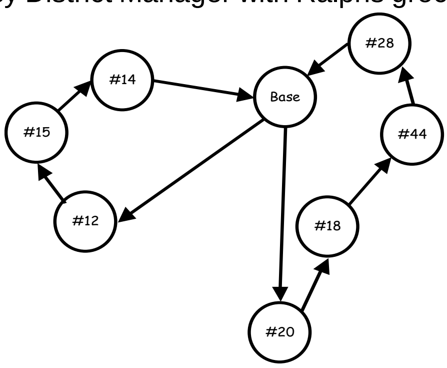


Figure 10*. Delivery Problem with 2 Loops*

The only problem with this occurs if the loops are very unbalanced. To keep this from happening, a constraint is added that determines the minimum number of stops within each loop in **D18**, which is determined by the formula

D18: =**MIN**(**ABS**(**D16-D17**)-1,**ABS**(**MIN**(**D16,D17**)+**MAX**(**E4:E14**)-**MAX**(**D16,D17**)))

In this case, **D16** is store 1, **D17** is store 2, and we take the minimum of (2-1) and (10-1) computes the number of stops between the two Oaklands and wants to be at least 5 to even out the two routes at 5 and 6. So the starting setup is not feasible for our purposes yet and shown in **FIGURE 11**.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	PDM Northern California Delivery Problem									MPH	Audit	min/hr					
2	PDM			2	# Routes					45	25	60					
3		From	To	Miles		**Minutes**	Oakland1	Oakland2	Folsom	Granite Bay	Rancho Cordova	Rosev ille	Sac (966)	Sac (970)	Sac (979)	San Ramon	Santa Rosa
4	1	Oakland1	Oakland2	999	1	Oakland1	999	999	112	110	105	110	100	97	100	21	70
5	2	Oakland2	Folsom	112	2	Oakland2	999	999	112	110	105	110	100	97	100	21	70
6	3	Folsom	Granite Bay	10	3	Folsom	112	112	999	10	10	15	14	25	22	108	127
7	4	Granite Bay	Rancho Cordova	11	4	Granite Bay	110	110	10	999	11	7	17	30	30	105	125
8	5	Rancho Cordova	Roseville	12	5	Rancho Cordova	105	105	10	11	999	12	7	20	15	101	120
9	6	Roseville	Sac (966)	20	6	Roseville	110	110	15	7	12	999	20	26	28	105	124
10	7	Sac (966)	Sac (970)	15	7	Sac (966)	100	100	14	17	7	20	999	15	9	95	114
11	8	Sac (970)	Sac (979)	3	8	Sac (970)	97	97	25	30	20	26	15	999	3	92	112
12	9	Sac (979)	San Ramon	95	9	Sac (979)	100	100	22	30	15	28	9	3	999	95	115
13	10	San Ramon	Santa Rosa	85	10	San Ramon	21	21	108	105	101	105	95	92	95	999	85
14	11	Santa Rosa	Oakland1	70	11	Santa Rosa	70	70	127	125	120	124	114	112	115	101	999
15				1432													
16	#	Index	Position	#stops													
17	1	1	1	0		Min ≥	Stops										
18	2	2	2	9		0	4										

Figure 11. *Setup for Two Loops*

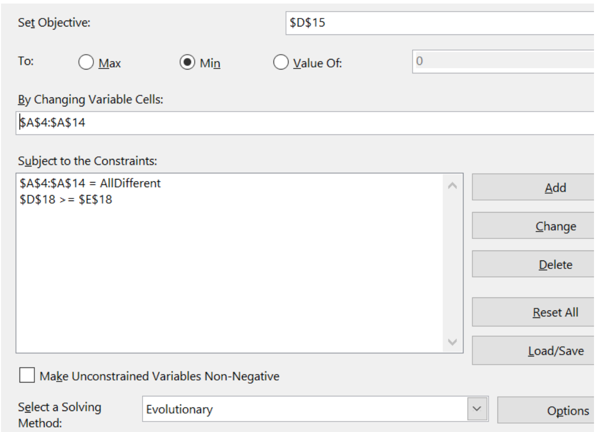


Figure 12. *Solver Setup*

Cells **A4:A14** contain the Alldifferent variables and it's not necessary to guess the answer, so numbering them in order will do. Notice that this starting solution is not feasible, as Oakland1 goes to Oakland 2 immediately and that there would be a loop of 1 city and 10 cities in rows 16-18.

The simple Solver setup is shown in **FIGURE 12**.

The Solution is shown in **FIGURE 13** and the Report in **FIGURE 14**.

The equations to force somewhat equal tours is given in rows 17-18.

A17: 1 **B17**: =**VLOOKUP**(A17,$A$4:$E$39,5,0) **C17**: =**SMALL**(B17:B18,A17) **D17**: =C18-C17-1

A18: 2 **B18**: =**VLOOKUP**(A18,$A$4:$E$39,5,0) **C18**: =**SMALL**(B17:B18,A18)

D18: =**MAX**(A4:A14)-**SUM**(D17:D17)-D2 **F18**: =**MIN**(D17:D18) **G18**: 4

Hesse, Rick Excel Student Routing Projects

	A	B	C	D	E	F	G
3		From	To	Miles		**Minutes**	Oakland1
4	9	Sac (979)	Sac (970)	3	1	Oakland1	999
5	8	Sac (970)	Santa Rosa	112	2	Oakland2	999
6	11	Santa Rosa	Oakland2	70	3	Folsom	112
7	2	Oakland2	Roseville	110	4	Granite Bay	110
8	6	Roseville	Granite Bay	7	5	Rancho Cordova	105
9	4	Granite Bay	Folsom	10	6	Roseville	110
10	3	Folsom	Rancho Cordova	10	7	Sac (966)	100
11	5	Rancho Cordova	Sac (966)	7	8	Sac (970)	97
12	7	Sac (966)	Oakland1	100	9	Sac (979)	100
13	1	Oakland1	San Ramon	21	10	San Ramon	21
14	10	San Ramon	Sac (979)	95	11	Santa Rosa	70
15				545			
16	#	Index	Position	#stops	Miles		
17	1	10	4	5	244	Min ≥ Stops	
18	2	4	10	4	301	4	4

***Figure 13**. Solution for Two Tours*

Delivery Report Route #1

From	To	Miles	Time Travel	Time Store	Minutes	Hours
Oakland1	Santa Rosa	70	93.3	25		
Santa Rosa	Sac (970)	112	149.3	25		
Sac (970)	Sac (979)	3	4.0	25		
Sac (979)	San Ramon	95	126.7	25		
San Ramon	Oakland2	21	28.0			
		301	401.3	100	501.33	8:21:20

Delivery Report Route #2

From	To	Miles	Time Travel	Time Store	Minutes	Hours
Oakland2	Roseville	110	146.7	25		
Roseville	Granite Bay	7	9.3	25		
Granite Bay	Folsom	10	13.3	25		
Folsom	Rancho Cordova	10	13.3	25		
Rancho Cordova	Sac (966)	7	9.3	25		
Sac (966)	Oakland1	100	133.3			
		244	325	125	450.33	7:30:20

Totals Miles	Time Travel	Time Store	Minutes	Hours
545	727	225	951.67	15:51:40

***Figure 14.** Tour Report*

The Report shows two tours, both starting from Oakland, for two days of travel close to 8 hours each day. This will allow Reda to land in Oakland at night and visit the pharmacies over the next two days and take an evening flight back to Los Angeles. This determines the DP with two days for routes to visit all stores.

DP#4 (Teemu) Numero Uno Pizza

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	NUMERO UNO PIZZA VISITS														
2	NUP405a			4											
3	#	Store	Store	Miles	Position		Miles	Office	Office	Office	Office	3	5	6	8
4	26	72	Office	1	1		Office	999	999	999	999	16	16	9	8
5	2	Office	28	2	2		Office	999	999	999	999	16	16	9	8
6	18	28	71	5	3		Office	999	999	999	999	16	16	9	8
7	25	71	6	2	4		Office	999	999	999	999	16	16	9	8
8	7	6	57	4	5		3	16	16	16	16	999	6	15	9
9	23	57	35	12	6		5	16	16	16	16	6	999	13	10
10	20	35	401	30	7		6	9	9	9	9	15	13	999	18
11	28	401	10	21	8		8	8	8	8	8	9	10	18	999
12	10	10	29	24	9		9	21	21	21	21	29	29	29	20
13	19	29	Office	28	10		10	51	51	51	51	63	64	56	57
14	3	Office	47	12	11		11	18	18	18	18	11	13	22	13
15	22	47	3	3	12		16	19	19	19	19	26	27	26	17
16	5	3	43	7	13		17	9	9	9	9	15	15	21	6
17	21	43	21	3	14		21	22	22	22	22	7	8	19	18
18	14	21	5	8	15		22	37	37	37	37	24	27	36	33
19	6	5	Office	16	16		26	23	23	23	23	8	8	19	18
20	4	Office	8	8	17		27	27	27	27	27	17	12	23	22
21	8	8	11	13	18		28	2	2	2	2	17	17	7	9
22	11	11	74	19	19		29	28	28	28	28	41	41	33	34
23	27	74	22	27	20		35	21	21	21	21	27	25	15	29
24	15	22	27	20	21		43	23	23	23	23	7	10	19	18
25	17	27	26	7	22		47	12	12	12	12	3	11	16	8
26	16	26	Office	23	23		57	7	7	7	7	19	17	4	18
27	1	Office	59	4	24		59	4							
28	24	59	16	17	25		71	7							
29	12	16	9	4	26		72	1							
30	9	9	17	15	27		74	31							
31	13	17	72	9	28		401	41							
32		Total DP Miles		344											
33															
34	#	Index	Position	# stops	Miles	Time									
35	1	24	2	8	128	11:39	Min ≥	Stop							
36	2	2	11	5	49	6:24	5	5							
37	3	11	17	6	117	9:20									
38	4	17	24	5	50	6:25									
39	1				344										

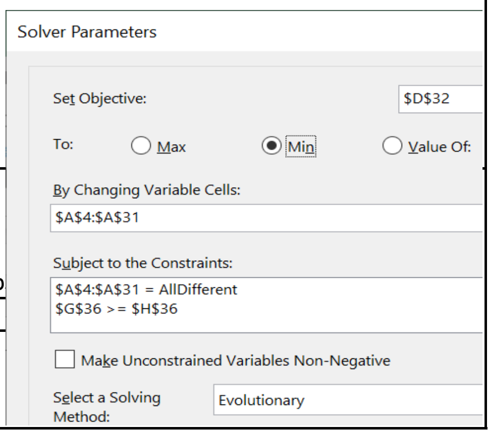


Figure 15. Solution for DP#4 Numero Uno Pizza Visits

Referring to Teemu Kivviku and his **TST** for Numero Uno Pizza, Teemu also needed to visit each store once a month to check on the accounts and keep in touch with each franchisee. He also developed a Delivery Problem, in this case DP-4 so he could visit all 28 stores during the first week of each month on Monday, Tuesday, Wednesday and Thursday and take Friday off. **Figure 15** shows the **DP#4** template for four (4) loops and the first attempt was to require at least 5 stores, allowing for maybe 6 or 7 to even out to 6 per day.

Pizza Tour - 4 Loops, 24 stops

# stops	Miles	Time	**Solution 1**	
8	128	11:39	Min >	Stops
5	49	6:24	5	5
6	117	9:20		
5	50	6:25		
	344			

# stops	Miles	Time	**Solution 2**	
6	155	10:25	Min >	Stops
6	116	9:18	6	6
6	61	7:44		
6	43	7:13		
	375			

# stops	Miles	Time	**Solution 4**	
6	54	7:32	Min >	Stoos
6	64	7:49	6	6
6	123	9:30		
6	114	9:15		
	355			

# stops	Miles	Time	**Solution 4**	
6	32	6:54	Min >	Stops
6	107	9:03	5	5
5	124	8:32	Max <	Stops
7	94	9:41	7	7
	357			

Figure 15 shows the miles and times have to be manually added and the blue cells in column E mark the beginning of each tour and the number of miles is gathered from column D and the time to visit at each stop is 60 minutes times number of stores and the time is takes to drive the miles at 35 MPH. Note that this solution is quite unblanced for 344 miles, but the miles are indeed minimized. However, what is crucial in this case is the number of stores visited each day at one hour each, not the mileage.

Figure 16 shows 4 different solutions to this DP#4.

One important problem occurs with having one very long tour lasting almost 12 hours, and the shortest just 6 ½ hours. To deal with this, a second constraint can be added in row 38 to put an upper limit on the number of stores for each loop or tour. Shown in **Figure 16** is the summary of 4 different trials and in this situation, mileage is not nearly as important as balancing time.

Figure 16. *Four Different Solutions*

Run 1 solution is also shown again. **Run 2** attempts to even out the number of stores at 6 each, but the time varies by over 3 hours from shortest (7:13) to longest (10:25), a difference of 3 hours, 12 minutes. Again miles are not that important. **Run 3** shows an equal number of stops and only 11 miles more than the best.

Run 4 gets the variation expected with >5, but the upper bound helped change solution in Run 1 to a more equitable one that Teemu could use. The difference between shortest and longest time is 2 hours, 47 minutes and only 13 miles longer than Run 1.

NBA Auditor's Tour DP#9: (Kurtis)

Kurtis Campbell, a MS in Analytics student at LMU from Montreal, Canada, was angling for an analytics position with the Phoenix Suns of the NBA. He hypothetically assumed an audit team of accountants/analysts from the NBA Office in New York had to visit each team at their city headquarters. But temporarily placing this analysis team to work from Atlanta would be advantageous to cut down on travel time/cost more in the middle of the country.

Given 9 tours, determine a good travel schedule. This was to be a much larger problem in terms of number of cities and needing 9 tours. There are 30 NBA teams but only 28 cities, since New

York Knicks and Brooklyn Nets (NY/BYK) are only 4 miles apart and the Los Angeles Lakers and Los Angeles Clippers occupy the same arena at which they each have their headquarters (Staples Center). And if the analytics team situates in Atlanta, the Hawks can be visited at any time and thus reduces the number of cities to be visited to 27. **Figure 17** shows part of the template, with the first row and column duplicated 8 more times for a DP#9 AllDifferent solution.

NBA Commisioner Tour 9 Trips

IND Project | 9 | Number of Trips

#	City	City	Miles	Position
20	NO	ATL1	424	1
1	ATL1	DET	626	2
15	DET	TOR	198	3
23	TOR	CLE	182	4
14	CLE	ATL9	572	5
9	ATL9	ORD	389	6
21	ORD	MIA	219	7
17	MIA	ATL5	607	8
5	ATL5	MIL	716	9
18	MIL	MIN	392	10
31	MIN	POR	1,350	11

Team	ATL1	ATL2	ATL3	ATL4	ATL5	ATL6	ATL7	ATL8	ATL9	BOS	BKY
ATL1	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL2	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL3	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL4	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL5	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL6	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL7	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL8	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
ATL9	9999	9999	9999	9999	9999	9999	9999	9999	9999	957	754
BOS	957	957	957	957	957	957	957	957	957	9999	204
BKY	754	754	754	754	754	754	754	754	754	204	9999

***Figure 17**. DP#9 for NBA 37 Cities*

Total Tour Length 19,050

#	Position	# stops
1	2	3
2	31	2
3	25	5
4	15	4
5	9	4
6	28	2
7	36	2
8	20	4
9	6	2
1		

Min ≥	Equal
2	2
Max ≤	**Equal**
5	5

This shows one run of DP#9 with lower (2) and upper limits (5) on number of cities for the 9 loops and a total distance of 19,050 miles based in Atlanta.

***Figure 10.** 9 Tours for NBA Visits*

Given the size of the tour in terms of cities and number of loops or tours, it is not surprising that we might need both lower and upper bounds to get a solution which is sensible. Minimizing miles is important but not critical, so look at **Figure 17** for the results of three different runs. In fact given the number of variables, trying different starting values with the same constraints yields different numbers and tours to consider.

Summary DP#9 NBA 27 cities, 9 tours									
					Min > Cities				
Run 1		Total	17,736		2	2			
# cities	2	2	2	3	2	3	9	2	2
Miles	1773	1496	1215	1322	1134	2159	5269	1453	1915
					Min ≥ Cities				
Run 2		Total	24,162		3	3			
# cities	3	3	3	3	3	3	3	3	3
Miles	2131	1359	4791	1722	4621	2237	4092	1701	1508
					Min ≥ Cities			Max ≤ Cities	
Run 3		Total	20,167		2	2		5	5
# cities	3	2	3	2	5	3	2	5	2
Miles	1465	1134	1849	1508	5185	2210	1079	4522	1215
					Min ≥ Cities			Max ≤ Cities	
Run 4		Total	20,682		2	2		4	4
# cities	4	2	3	4	2	3	4	3	2
Miles	5130	1215	1435	1695	1915	1809	4453	1710	1320

Figure 19.
Summary of NBA DP#9 Tours

Run 1 has just a lower bound of 2 cities, and yields 6 tours of 2 cities, 2 tours of 3 cities and one very long tour of 9 cities for over 5,000 air miles.

Run 2 requires that all 9 tours be 3 cities and the total miles increases by over 6000 miles or 36.2%. Starting the template with the initial AllDifferent values as 1-2-3-4----36 yielded an infeasible solution, so the initial starting point was 1-10-11-12-2-13-14-15-3-16-17-18 and so on, using a feasible solution of 40,869 miles as a starting point. I was surprised that it just did not freeze at the start, but it did find a much better solution.

These two runs are two extremes.

Run 3 allows between 2 and 5 cities per tour and reduces Run 2 by almost 4,000 miles and required an upper and lower bound for # of stops.

Finally **Run 4** allows only tours of 2, 3 or 4 cities and is about 500 miles more than Run 3 by eliminating tours of 5 cities and also required an upper and lower bound for # of stops. The Report for this run is shown in **Figure 20**.

Of course mileage would be far greater if the teams had to use New York as their base, which may be why the NBA Commissioner would require teams to come to his office rather than have teams visit on site. But this scenario did present some interesting dynamics and insights into using the AllDifferent constraint.

Trip 1			Trip 2			Trip 3		
From	To	Miles	From	To	Miles	From	To	Miles
ATL7	MIA	607	ATL6	BKY-NY	754	ATL8	WSH	538
MIA	ORD	219	BKY-NY	BOS	204	WSH	PHIL	122
ORD	ATL5	389	BOS	ATL3	957	PHIL	ATL1	660
Total		1,215	Total		1,915	Total		1,320

Trip 4			Trip 5			Trip 6		
From	To	Miles	From	To	Miles	From	To	Miles
ATL5	CHI	580	ATL3	NO	424	ATL2	DAL	725
CHI	MIL	138	NO	HST	318	DAL	OKC	211
MIL	IND	291	HST	SA	188	OKC	MEM	422
IND	ATL4	426	SA	ATL9	879	MEM	ATL8	352
Total		1,435	Total		1,809	Total		1,710

Trip 7			Trip 8			Trip 9		
From	To	Miles	From	To	Miles	From	To	Miles
ATL1	MIN	1050	ATL4	CTL	245	ATL9	LAC-L	1942
MIN	POR	1,350	CTL	DET	498	LAC-L	GSW	313
POR	UTH	648	DET	TOR	198	GSW	SAC	85
UTH	PHX	483	TOR	CLE	182	SAC	DEN	893
PHX	ATL7	1,599	CLE	ATL6	572	DEN	ATL2	1,220
Total		5,130	Total		1,695	Total		4,453

TOTAL	20,682	7,780			5,419			7,483

Figure 20*. Report for NBA Run 4*

<u>Kurtis TST Commissioner's Tour of all 30 teams from NYC</u>

Kurtis also constructed a large template to determine the optimal TST of all 30 teams in the NBA by the Commissioner from New York City. The two teams in New York will be treated separately (NYK, BKLY), while Los Angeles Lakers and Clippers will be treated as one, with both teams headquartered in the same building. There are thus 30 stops to make and the best minimum travel miles so far is 9,884 miles.

Hesse, Rick — Excel Student Routing Projects

REPORT

From	To	Travel Miles
NY	PHIL	92
PHIL	WSH	122
WSH	CTL	296
CTL	ATL	245
ATL	MEM	352
MEM	OKC	422
OKC	DAL	211
DAL	HST	224
HST	NO	318
NO	ORD	533
ORD	MIA	219
MIA	SA	1158
SA	PHX	849
PHX	LA	362
LA	GSW	313
GSW	SAC	85
SAC	POR	487
POR	UTH	648
UTH	DEN	374
IND	CHI	158
DEN	MIN	712
CHI	MIL	138
MIN	MIL	392
MIL	DET	257
DET	CLE	98
CLE	TOR	182
TOR	BOS	429
BOS	BKY	204
BKY	NY	4
Total Miles		9884

<u>The rest of the story</u>

Teemu Kivviko is now Sales Director at Metsä Board Americas Corporation of Finland and Logistec in New York city and been with the company since 2003. The company makes sustainable cardboard packaging utilizing huge container ships for cardboard and paper products across the Atlantic.

Kurtis Campbell was on the LMU golf team as an undergraduate, earned his PGA card and was a graduate assistant at LMU working with basketball analytics, which led to a short stint with the Phoenix Suns online. He got his dream job and location being employed by Lightspeed based in his hometown of Montreal. He is working with an analytics division that services country clubs with golf analytics. Kurtis also took my statistics/forecasting course, and for his project did a very nice analysis using conditional probability to determine at what distance does a golfer decide to go straight for the hole, or two-putt the hole.

Reda Alhoussayni did accept the job with Ralph’s Groceries and at one time managed 70 different locations. He now is President of his own two pharmacies, Olive Branch Pharmacies, in Los Angeles since 2009. Reda is also an Adjunct Assistant Professor at USC (where he obtained his Pharmacy degree) and uses his pharmacies for internships for USC Pharmacy students.

CONCLUSION

The application of using the AllDifferent constraint in the Excel Solver to routing has opened up practical applications for several of my MBA students over the years. The Traveling Sales Tour (**TST**), Traveling Sales String (**TSS**) and Delivery Problem (**DP#n**) with **n loops** occur in many business situations. My own experience and that of my students with these heuristic algorithms in Excel leads me to mention a few important points about these hueristics.

1. Try different starting values and see if the resulting answers change. I've even used all 1's before and let the computer randomly select 1➔n as a starting point.

2. The Solver has many stopping points built into it, so keep pressing <Enter> to keep going until you've had enough. You can then always press <Esc> and the Solver will ask you if you want to stop at that point. It will save the best solution for that run.

3. There is no guarantee that any of the answers are optimal unless you have a way of solving it optimally in another manner. Remember that you are looking for ia good and useable (feasible) one, not necessarily optimal.

4. Be careful not to accept a bad solution – so pay attention to the details of the solution(s) you choose. Kurtis initially handed in a solution that have several 2-city tours because he did not add minimum constraints for the last 3 tours, and saw only that the total mileage was minimized but created very long loops and several short ones that would be impractical.

5. My students consistently expressed their highest appreciation for the experience of applying the templates which are covered in the course to their own situation, and specifically their term-end project for Data Analytics majors. They may not have been happy about the hard work in the course, but they certainly could see the applications.

The Excel templates are loaded on the DSI Conference 2021 web site and the four Excel files are named Kurtis, Reda, Rick and Teemu which cover all the files for **TST, TSS** and **DP-n** examples given in this paper.

REFERENCES

ALLDIFFERENT constraint in Solver,
Retrieved
https://www.google.com/search?q=ALLDIFFERENT+constraint+in+Solver&rlz=1C1CHBD_enUS773US773&oq=ALLDIFFERENT+constraint+in+Solver&aqs=chrome..69i57j0i22i30.11708j0j7&sourceid=chrome&ie=UTF-8

Alhoussayni, R, LinkedIn, https://www.linkedin.com/in/reda-alhoussayni-a3413437/

Campbell, K, LinkedIn, https://www.linkedin.com/in/kurtis-campbell-18aa4a15a/

Kivikko, T, LinkedIn, https://www.linkedin.com/in/teemukivikko/

Hesse, R., LUMAT4, Chapter 405 Integer Programming, pgs 1-20, Blue Pond Publishing (AKA Azel Publishing), eTextbook, 1998-2021.

Hesse, R., "Alldifferent Constraint and the Traveling Sales Problem Heuristics – Part I," *Decision Line*, Vol 36, No 3, September 2005, pages 10-13.

Hesse, R., "Alldifferent Constraint and the Traveling Sales Problem Heuristics – Part II," *Decision Line*, Vol 36, No 4, November 2005, pages 7-10.

Hesse, R., Weinstock, J, Grum, A.F., "Term-end Scheduling at United States Military Academy", Conference Presentation, Southeast Decision Sciences, March 1982, Williamsburg, VA.

DECISION SCIENCES INSTITUTE

Queue Modeling for Decision Support at a Henkel Call Center

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

Yuzhen Liang
Boston University Metropolitan College
Email: yuzhenl@bu.edu

David Cadreact
Boston University Metropolitan College
Email: dcadreac@bu.edu

Mingxuan Zhang
Boston University Metropolitan College
Email: zmxming@bu.edu

Alexandra Malaspina
Henkel Corporation
Email: alexandra.malaspina@henkel.com

ABSTRACT

With shorter lifespans and increased complexity of products and services, call centers play an increasingly important role in many businesses. At Henkel, B2B customers and consumers contact help desks for many purposes, including getting assistance with the application of new adhesive products. Modeling the call center as a queuing system required the thorough analysis of its unique structure. A discrete-event simulation was developed that, in turn, generated a metamodel of optimization parameters. A prototype decision support system was created that is used by the call center manager for personnel planning.

KEYWORDS: Call center, Simulation, Metamodeling, Queuing, Decision Support

INTRODUCTION

Call centers play an important role in many businesses, due to the shorter lifespan and increased complexity of products being sold. In fact, the revenue generated by the call center industry sector was $25.1 billion in 2020, and it is forecasted to increase 2.3% per year (IBISWorld, 2020). These numbers do not include the various forms of call centers, help desks, support groups, or related customer or employee assistance operations that play a role in many firms. In this article, these operations will all be referred to as a call center. Because labor represents the most important resource at a call center, a manager should pay close attention to controlling labor costs. However, the nonlinear “hockey stick” relationship between labor utilization and customer waiting times makes maximizing labor utilization a poor planning strategy. The optimal utilization level should depend on many factors, including labor cost and the cost of waiting, which (although difficult to quantify) is lower for some businesses and higher for others. Hence, a decision support tool needs to accurate represents the call center's operations while accounting for both customer desires and company needs.

This article is the culmination of a project performed at a call center operated by Henkel. The call center is unique because it serves mainly business-to-business (B2B) customers, who use its adhesives as a requirement for their manufacturing processes. Although revenue is not directly generated by the call center, it plays a critical role in Henkel's business model, which focuses on services to customers. Therefore, the call center manager needs to assign personnel to the call center so that labor costs are controlled, while providing prompt customer service. A call center decision support system (DSS) will need to represent the queueing system accurately, with assumptions that conform to all characteristics that affect its operation.

This article is organized as follows. First, relevant literature that forms the basis of the work presented is discussed. Second, the methodology used to integrate the data analysis, modeling, and decision support system (DSS) is described. Third, the call center is described, focusing on its unique characteristics, followed by an analysis of data obtained from its operation. Fourth, the two simulations that were created to represent the call center's operation are described. Fifth, the results of running the simulations are presented with a focus on optimizing capacity utilization. Sixth, we describe the development and use of the prototype DSS at Henkel. Finally, future plans are discussed.

LITERATURE REVIEW

The importance of call centers for supporting successful business operations has been well established in the academic literature. For many firms, a call center serves as the primary form of customer interaction; it constitutes a large and growing component of the global economy (Aksin et al., 2007). Besides traditional business-to-consumer environments, call centers are also used for emergency dispatching (Sariyer, 2018), human resources support (Grobbelaar et al., 2004), and in healthcare (Liao et al., 2011). When effective, a call center will strengthen the firm's reputation; however, poor call center service is likely to result in customer defections (Dudin et al., 2013).

A direct relationship between call center performance, customer loyalty, and long-term profits has been reported (Gong et al., 2015). For example, it has been shown that, as waiting time increases, customer satisfaction decreases (Garcia et al., 2012; Peevers et al., 2009; Davis & Volmann, 1990; Bielen & Demoulin, 2007). Call centers can directly enhance a company's revenues through cross-selling, resulting in better customer retention (Örmeci & Aksin, 2010). Call centers play an important role in establishing a long-lasting relationship with the customer, thus increasing customer loyalty (Amoako et al., 2016; Gronholdt et al., 2000; Lia et al., 2014). In turn, loyalty is a strong indicator of how customers act in the future by impacting the likelihood of customers re-engaging with the same firm (Edvardsson et al., 2000).

The task of a call center manager is to plan capacity in ways that maximize benefits and minimize costs through optimizing the allocation of labor, which has resulted in call center queue analysis as a popular research focus (Aksin et al., 2007; Aktekin & Ekin, 2016). Queueing models are used to determine how many servers (e.g., typically called representatives or agents) should be assigned during specific time periods. They can also be used to help anticipate training requirements for human resource staff (Gans et al., 2003). Uncertainties associated with demand and service times must be taken into account in the queuing model because of its effect on service levels (Munoz & Bastian, 2016). In addition, planning for future product releases or feature updates adds to the challenge of training and staffing call center staff (Aksin & Harker, 2003).

Decision support for capacity planning in a call center can include several decisions. They range from short-term shift scheduling (Harrison & Zeevi, 2005) to long-term resource planning (Hasija et al., 2010). Effectiveness of decision support must account for system complexities, although certain assumptions are often present, such as Poisson arrivals where customers arrive independently of one another at a constant rate (Sariyer, 2018). At times, arrival rates change across days or hours and therefore the decision model should be adjusted accordingly, including the ability to recognize when demand patterns change (Gans et al., 2015). Although a simple queuing model often assumes exponential service times, other distributions are typically present in actual service systems. For example, the gamma distribution was shown to represent service times for various healthcare services (Millhiser & Veral, 2019).

For many firms, the most important visible factor for call center management is labor cost; a DSS needs to be especially accurate in this regard (Chromy et al., 2012). But the task of reconciling cost efficiency and customer satisfaction is challenging (Labach, 2010). Several queueing models have been developed to mimic various call center operations in attempts to solve the staffing problem. Dudin et al. (2013) model a call center with a call-back for customers who abandon the queue by considering two types of customers and two types of servers. Gong et al. (2015) model a call center where customers generate revenue and arrival rates are affected by customer satisfaction levels. Liao et al. (2011) model a call center where agents handle both inbound calls and “back of office” tasks such as answering emails and voicemails. Kanavetas and Balcıoǧlu (2018) developed a call center queueing model where queue length and wait time are visible to the customers who behave in response to this information. Other researchers have modeled call centers that provide better service to certain classes of the customer (Kim et al., 2013; Kim et al., 2016).

The study of customer contact centers has been addressed by analytics researchers for almost 100 years (Mehrotra & Fama, 2003). Over time, new technologies have affected call center operations as well as their analysis. David (2016) addressed the challenge of making call center personnel decisions without a proper simulation model to account for system complexity. Because information technology plays a key role in both its operation and its analysis, computing power and programming capability limitations can compromise an organization’s ability to effectively analyze its call center operations (Doomun, 2008). To be effective, the gap between the call center operation in reality and the simulation model needs to be minimized (Ma et al., 2011).

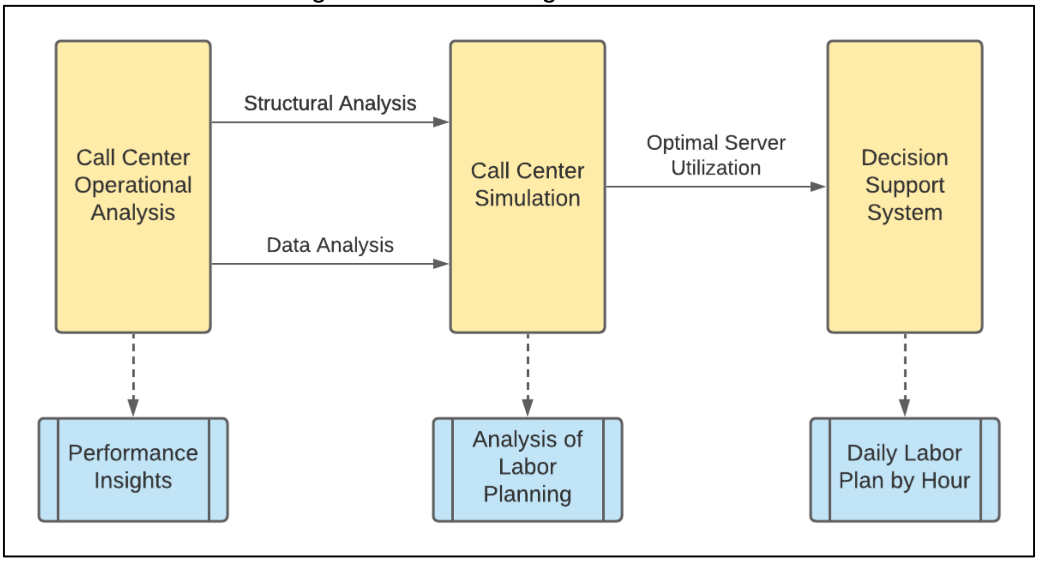
The Bell Canada call center provides an example of the use of simulation that can be adjusted to account for real-time changes in arrival rates, service times, and abandonment of customers (Avramidis & L'Ecuyer, 2005). The use of a simulation model was also illustrated by Cezik and L'Ecuyer (2008) who use a simulation model to explore flexible personnel scheduling frameworks. Another application of simulation modeling studied the integration of key division variables and outcomes to help the organization understand the call center’s impact on a firm (Mehrotra & Fama, 2003). The simulation of any operation requires many iterations that can take considerable computing time, which often needs to be run for many hours (Bouzada, 2009). The simulation approach may not be practical as a component of a DSS, although a metamodel derived from a simulation can be useful (Barton, 2015).

METHODOLOGY

Figure 1 provides a high-level framework for the development of the call center DSS. After becoming familiar with the call center's operation, information was collected over a six-month period. The information helped to determine nuances of the call center's operations and included data that were used to describe the statistical variation associated with call demand, agent service times, and caller abandonment characteristics. This information was used to create a representative model of the call center's operations and generate accurate simulation parameters.

The data analysis generated the input parameters for the simulations that were coded in Python. Two simulations were created and validated using queue length data available in the original data set. One simulation was designed for use by Henkel managers to experiment with effects of changes in their products, services, or operations. The other simulation was used to generate an optimal service utilization target using a customized mathematical approach. This approach was based on the hockey-stick waiting time versus server utilization relationship curve and inputs from the call center manager.

Figure 1: Methodological Framework



The optimal server utilization target became the basis for the prototype DSS, which was developed using a simple Excel framework. It requires information associated with all important features of the call center's operations, its callers, and its service agents. The system includes forecasts of caller arrival rates and adjustments to service times based on recent data. It will provide the call center manager with the option to optimize the labor allocation plan, or score the manager's plan while assisting with metrics for updating the plan.

CALL CENTER OPERATIONS

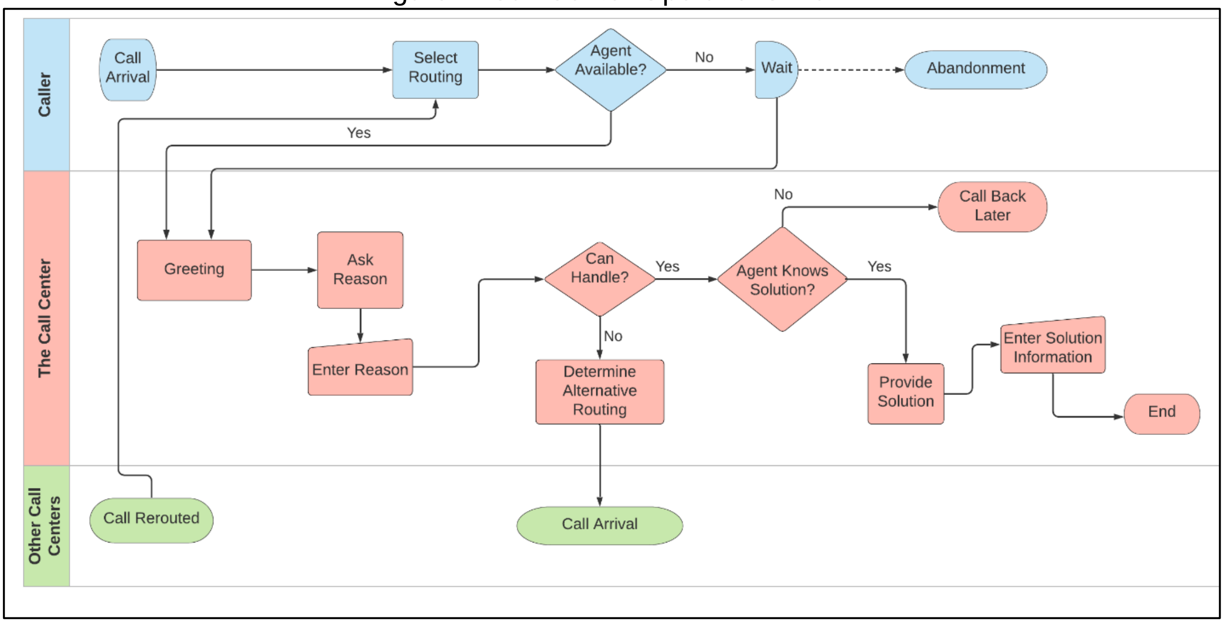
Henkel is a multi-national firm headquartered in Düsseldorf, Germany. As of 2021, Henkel employed about 53,000 people worldwide. It had sales in FY 2020 of 19.3 billion euros and operated in three markets: adhesive technologies, beauty care, and laundry & home care. Its adhesive division produces the Loctite brand and constitutes just under half of its total revenues.

Customers of Loctite adhesives include consumers and industrial customers. Industrial customers use Loctite adhesive to manufacture products in fields such as aerospace, automotive, and medical devices.

Henkel operates a number of call centers around the world. This article concerns one of their North American general industry technical support help lines. It will be referred to as "the call center" in the remainder of this article. The call center is mainly dedicated to adhesives customers – most callers represent industrial customers. Call center agents are skilled technicians and engineers who assist callers with applications, while being able to perform other tasks within the firm. Agents are not co-located and their assignments to the call center can be flexible over the course of a week or day. Most of the callers are current customers who seek answers regarding specific applications. In many cases, calls are motivated by product design or process changes where callers need assistance with making modifications to adhesive applications. Other callers are new customers and some callers are consumers.

Figure 2 describes the operation of the call center. Calls arrive either directly to the call center or from another Henkel call center. Some calls are routed from the call center to another Henkel call center. Agents who cannot address a question may obtain information from a caller and contact them later. Or, they may ask for assistance from a technician or engineer at the Henkel R&D facility. The call center operates Monday through Friday for 8 hours per day.

Figure 2: Call Center Operations Flow

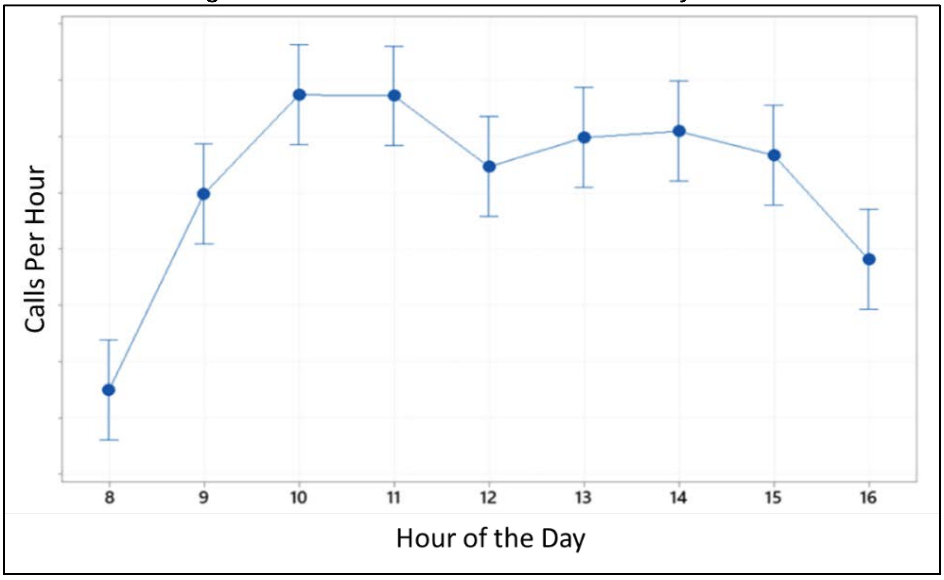


Call Center Data Analysis

Data were collected over a six month period in early 2020. The COVID-19 pandemic had affected the rate of calls but had no effect on the service times. The arrival patterns were analyzed first. Although not shown here, the arrival rates were stable across the period of data collection. There was no difference in arrival rates for each hour over the five operating days. However, as shown in Figure 3, the interval plot (i.e., the sample average and 95% confidence interval for the mean) shows that arrival rates varied by hour of the day. The interval plot's horizontal axis is the working hour (8 = 8:00 to 9:00 AM), and the vertical axis is the number of calls arriving. The vertical axis numbers are hidden to maintain confidentiality.

Next, the distribution of arrivals within each hour was determined. As illustrated in Figure 4, the arrival distribution was consistent with a Poisson distribution (this consistency was confirmed using goodness-of-fit tests). The Poisson distribution is common for services with random arrivals where customers act independently. Again, the display hides the horizontal axis numbers to preserve confidentiality. The display shows two bars for each time period – each shows the frequency - of the observed average and the expected average, both for the number of calls. In this case, only one representative time period is shown. Based on these results, the simulation will assume Poisson arrivals (i.e., the time between arrivals will follow an exponential distribution) and that the rate of arrivals will vary by hour.

Figure 3: Interval Plot of Arrival Rates by Hour



The analysis continues for service times. Although not shown, the service times for each agent were stable across the data collection period. The analysis of service times showed a difference across agents, as illustrated by the interval plot in Figure 5. In the display, the horizontal axis represents the agent (names are hidden) and the vertical axis represents the service duration. Again, the vertical axis numbers are hidden to maintain confidentiality. In the display, the length of the confidence intervals is inversely proportional to the number of calls for which data were analyzed.

Agent service times differ for two reasons: (1) their level of experience, and (2) their approach to customer engagement. For example, an agent with shorter call durations may be faster at getting a solution without much caller engagement, while another agent may provide similarly effective answers but spend more time engaging with the caller. The call center manager appreciates these differences and is not concerned as long as the agent provide accurate information to customers.

Figure 4: Comparison of Arrival Data and Poisson Distribution (Example)

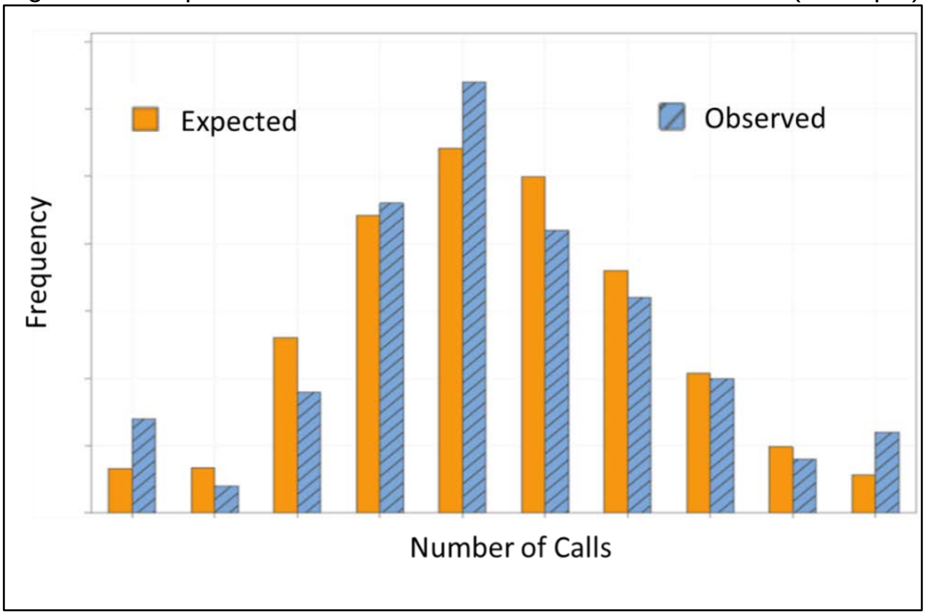
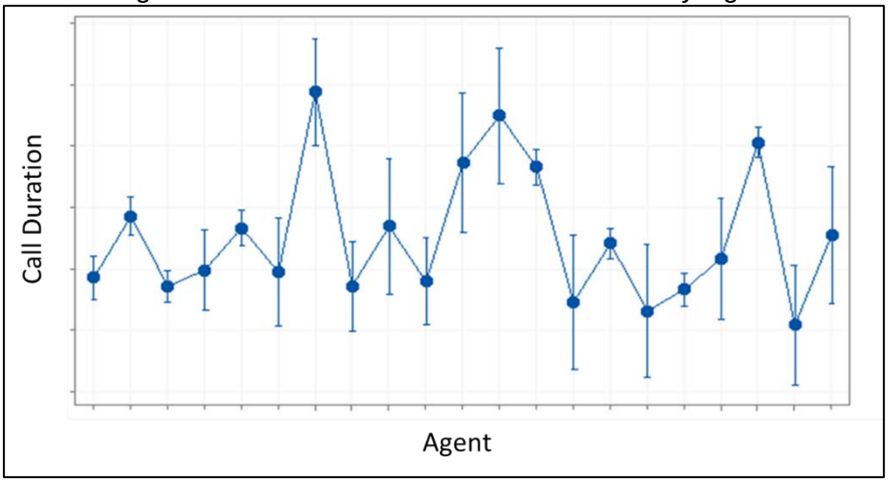
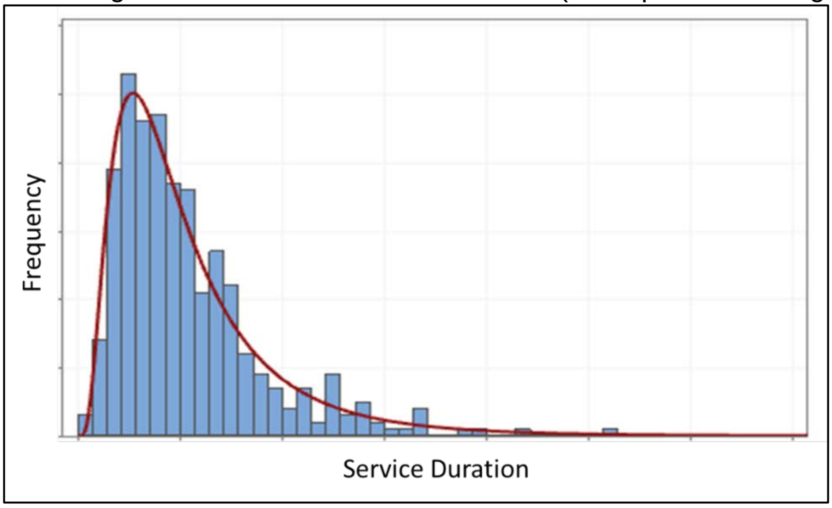


Figure 5: Interval Plot for Service Durations by Agent



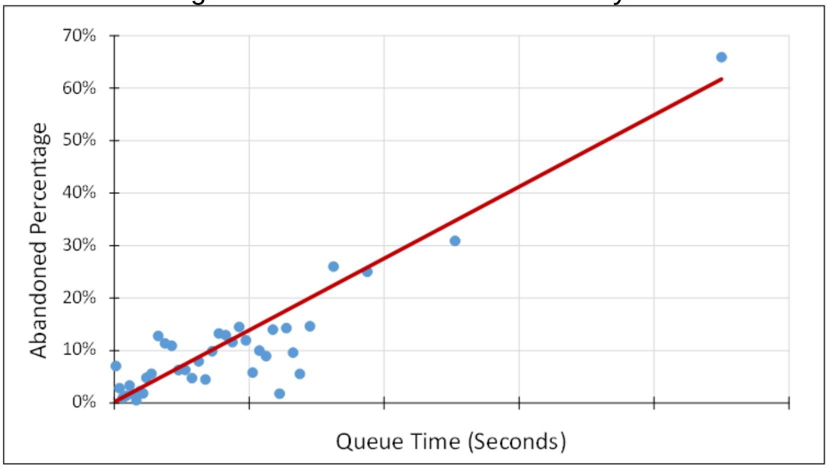
The distribution of service times for each agent was consistent with a log normal probability distribution, as illustrated in Figure 6 (which shows the call durations for one of the agents). The horizontal axis is service time and the vertical axis is frequency. The red curve is the log normal distribution with the parameters estimated based on the data set. Other agents showed similar patterns of variation, albeit with different mean values. The level of variation, measured as the coefficient of variation (CV), was consistently in the 60-80% range. Based on this analysis, the simulation will need to identify each agent by name to accurately represent service durations. This characteristic is uncommon for many call centers who strive to use standard work processes to lower call durations uniformly across agents.

Figure 6: Log Normal Service Time Distribution (Example for One Agent)



The analysis of caller abandonments (the occurrence of a caller entering a queue then leaving before service begins) proceeded as follows. Queue times were sorted and grouped. For each group, the average queue time was calculated, and the percentage of abandonments was tabulated. The results are shown in Figure 7, where the scatterplot shows the relationship between the queue time and the abandonment likelihood. The best fit was linear, as shown by the red line. The prediction when queue time is zero represents callers who left the call during or just after the automated welcome message.

Figure 7: Caller Abandonment Analysis



SIMULATION LOGIC

The simulation logic makes the following assumptions that are consistent with the call center's operation and the analysis of data described above. They are:

1. Arrivals following a Poisson process with a specified arrival rate (i.e., the simulation addresses one time period where the arrival rate is constant).
2. Callers wait in a common queue for service, and service is provided on a first-come first-served basis.

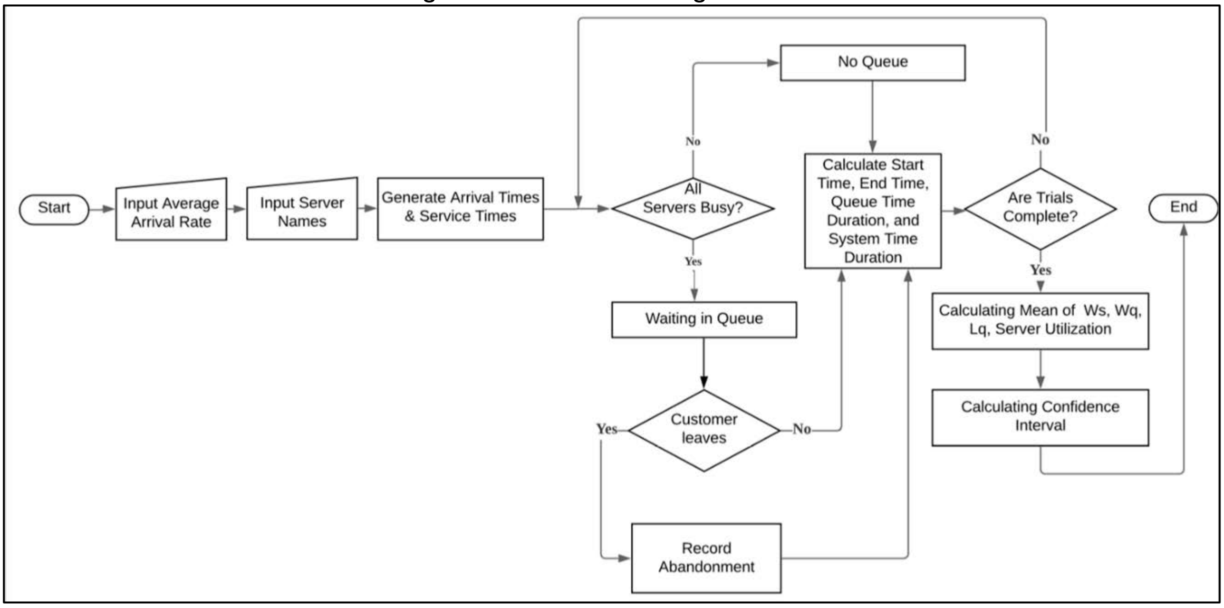
3. Before service begins, the likelihood that a caller in queue leaves the system is determined and some callers leave based on this likelihood.
4. Each caller is served by one agent.
5. Agents are chosen by name by the user and input to the simulation.
6. Service durations are log normal with a mean obtained from a database of each agent's average service time.
7. The variation of service durations has a CV of 0.75 (75%).
8. Callers exit the system upon completion of service.

The system's logic is described in Figure 8. The simulation model includes a database with agent names and their average service time. The following inputs are provided by the user: (a) the arrival rate (calls per hour), and (b) the identity of each agent who will provide service to callers (the maximum number is consistent with the current staffing levels at the call center).

Simulation outputs include the following:

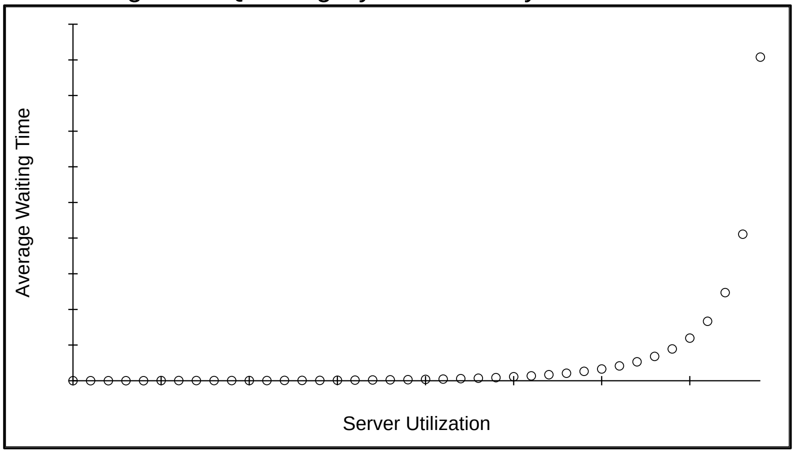
1. Estimated average waiting time in queue and 95% confidence interval.
2. Estimated median waiting time in queue.
3. Estimated average total time in system and 95% confidence interval.
4. Estimated average queue length and 95% confidence interval.
5. Estimate proportion of callers leaving the system before service and 95% confidence interval.
6. Server utilization.

Figure 8: Simulation Logical Flow



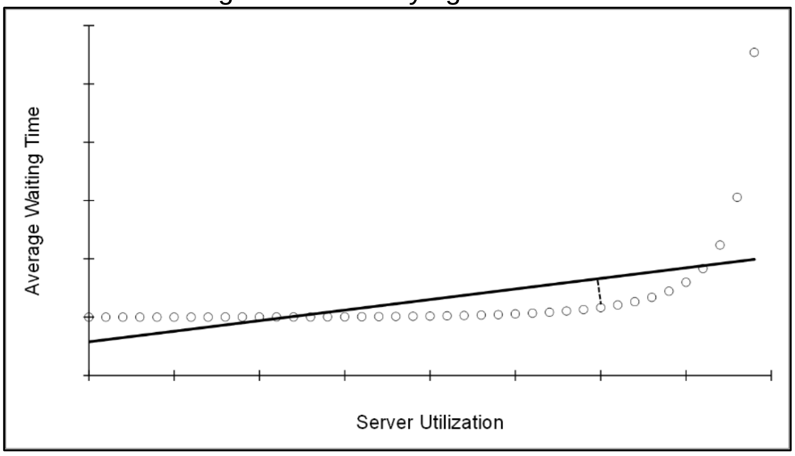
The approach for determining the optimal server utilization is based on the non-linear mathematical relationship between the server utilization and the average caller waiting time. This relationship will follow a so-called hockey-stick function (Figure 9). The hockey-stick pattern is found in any queuing system with either random arrivals or random service, and will vary in its parameters depending on the myriad of characteristics associated with the queuing system under study.

Figure 9: Queuing System Hockey-Stick Curve



The method chosen to determine the optimal server utilization for use in a capacity planning DSS is a variation of a method described by Salvador and Chen (2004). As illustrated in Figure 10, a linear regression line is fit to the hockey stick data, and the optimal server utilization is found at the mathematical "knee" of the curve. The knee is located at the longest Euclidean distance from the line to the curve.

Figure 10: Identifying the "Knee"



This method was chosen based on a parallel effort by other researchers who compared mathematical knee approaches to choices made by practitioners who were presented with hypothetical queue system hockey-stick curves. The method applied here provided the most precise consistency with practitioner choices. However, it needs to be calibrated based on the cost structure and customer expectations for each application. The calibration procedure is described later in this article.

SIMULATION RESULTS

The simulation models serve two purposes. The first simulation provides performance predictions for a user who inputs agent identities and the caller arrival rate. This simulation's logic was confirmed by comparing simulated results with actual call center performance data that were included in the original data set. During this confirmation exercise, it was noticed that the predicted server utilizations were low and the corresponding waiting times were very short. The call center manager confirmed that when demand slowed during the COVID-19 pandemic, call center staffing levels were maintained and the results were just as expected.

The second simulation determined the knee (optimal server utilization) using an iterative procedure. The user inputs the arrival rate and the number of servers, then the simulation code iterates across server utilizations from 50% to 99% in increments of 1%. By determining the average waiting time for each iteration, this procedure produces a hockey-stick curve. The knee is calculated using the approach described earlier. This code was used to generate knee values across a range of potential operational levels based on a structured experimental design.

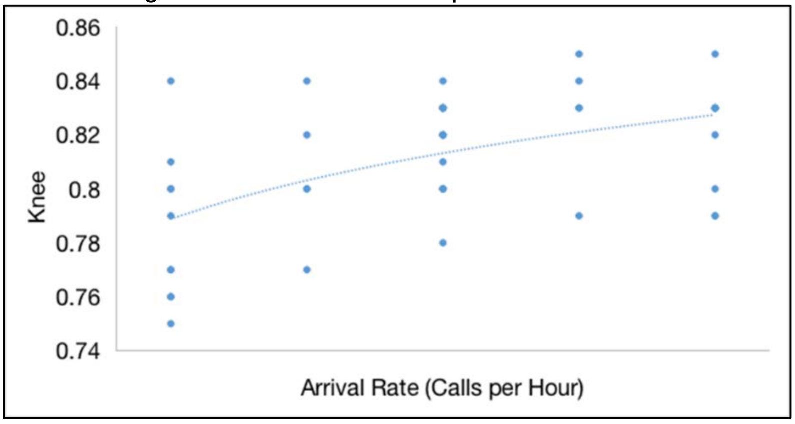
The second simulation was run across a full factorial experimental design with three arrival rates and three server numbers. For this analysis, all agents were assumed to operate with the same average service time (later, the first simulation code was used to verify that no differences in knee values would exist if the multiple servers had varying capacities). This analysis concluded that the number of servers did not influence the knee, but that the knee differed according to the caller arrival rate. No interaction effect of agents and arrival rate was present. As shown in Figure 11, the knee varied according to the arrival rate in a non-linear way. During these experiments, the service time CV was not adjusted because a prior simulation study concluded that service time CV did not affect a similar queueing system's performance when the CV exceeded 50% (Corlu et al., 2021).

Figure 11: Main Effect Plot for Arrival Rate vs. Knee

The second simulation model was run for intermediate arrival rates to create a metamodel for the knee based on the simulation results. The resulting metamodel is displayed in Figure 12. Equation 1 (where λ is the arrival rate) is the metamodel for the knee (rescaled to protect confidentiality of the data). The accuracy of this model was validated using the first simulation code. The simulation outcomes are expected to vary, and hence the dispersion of outcomes around the metamodel should not be troubling. This model generates the optimal server utilization but needs to be calibrated for specific applications, as detailed in the next section.

$$Knee = 0.5931 + 0.0734\ln(\lambda) \quad (1)$$

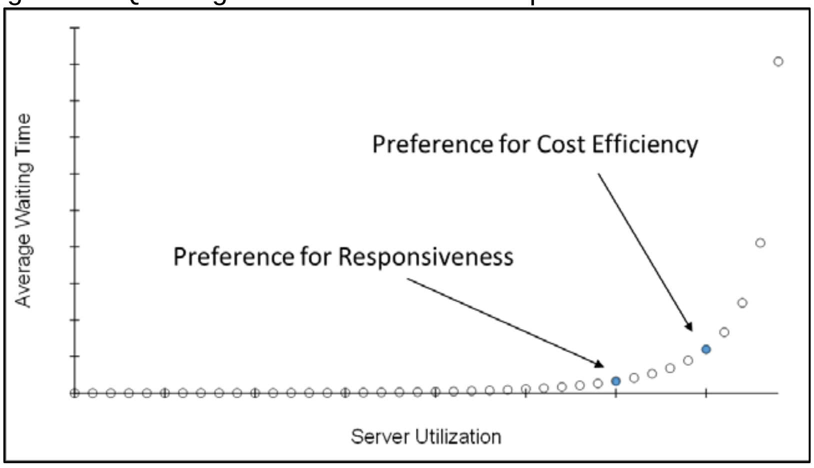
Figure 12: Metamodel Required for the DSS



PROTOTYPE DSS

The knee calibration procedure consisted of presenting the call center manager with a set of 9 hockey-stick choices, with each horizontal and vertical axis scaled to match the performance of the call center. The nine scenarios represented the range of agent assignments and arrival rates used in the experimental design referenced above. As shown in Figure 13, a call center manager would choose the optimal server level representing their most effective reconciliation of agent cost (i.e., server utilization) and caller promptness (i.e., average waiting time). A manager with lower server costs and/or a desire for prompt service would tend to choose the responsiveness preference, while managers with high server costs and/or less concern for promptness of service would tend to choose the cost efficiency preference. The calibration results for this call center were not shown to maintain confidentiality.

Figure 13: Queuing Model Calibration for Optimal Server Utilization



An example of the prototype DSS is shown in Figure 14. Although this system will be fine-tuned and improved going forward, the first version is designed for maximum user flexibility. When using the system, the user will enter the agent to be assigned during each hour (by simply adding a “1” to the assignment matrix). For each hour, the system determines the server utilization and computes its deviation from the optimal utilization, which was determined by the knee (Equation 1) and the calibration procedure. The average of the differences between the optimal utilization and user-generated utilizations is then calculated and converted to a green-yellow-red color. An example of the color code would be green if the average difference is less than 10%, red if the average difference is more than 15%, and yellow otherwise.

The manager would iteratively change the agent assignments based on a desire to achieve server utilization targets and other concerns they have as a manager of professional workers. The other concerns include the need for agents to attend training programs for part or all of a day, personal time off for agents, and special projects to which an agent may be assigned. The plan going forward is to ask the call center manager to use the porotype system for several months while documenting confusion, challenges, benefits, and desires for additional features.

CONCLUSIONS AND FUTURE WORK

This article detailed a comprehensive effort that used simulation and metamodeling to create a DSS for a Henkel call center. It required knowledge of the call center's operating structure and analysis of its operational and performance data. Two simulations were created that were integrated to generate inputs for the DSS and to validate the simulation and metamodel. It is presently being used by the call center manager to help plan weekly personnel assignments.

Figure 14: Prototype DSS

Call Center Capacity Planning System

Demand Rates:											
	Hour Starting:		**8:00**	**9:00**	**10:00**	**11:00**	**12:00**	**13:00**	**14:00**	**15:00**	
	Average Calls per Hour		8	13	17	20	32	30	26	22	
Target Server Utilization:			**77%**	**78%**	**79%**	**80%**	**81%**	**81%**	**81%**	**80%**	
Capacity Plan:											
Agent Name	**Average Service Time (Min)**	**Capacity (Calls per Hour)**	**8:00**	**9:00**	**10:00**	**11:00**	**12:00**	**13:00**	**14:00**	**15:00**	
Alice Jones	10.0	6.0	1	1	0	0	1	1	0	0	
George Miller	10.0	6.0	1	1	1	0	0	0	1	1	
Ivan Drago	9.0	6.7	0	0	0	1	1	1	0	0	
Laura Kelly	9.0	6.7	0	0	1	1	1	0	1	0	
Pat Vasquez	8.0	7.5	0	0	1	1	1	1	0	0	
Nancy Wu	8.0	7.5	0	0	0	1	0	1	1	1	
Philip Smith	7.0	8.6	0	0	0	0	0	0	1	1	
Mason Howard	7.0	8.6	0	0	0	0	1	1	1	1	
	Hourly Capacity (Calls/Hour)		12.00	12.00	20.17	28.33	35.40	36.24	37.31	30.64	
	Agent Utilization (%)		66.7%	108.3%	84.3%	70.6%	90.4%	82.8%	69.7%	71.8%	
	Deviation from Target		-9.9%	30.1%	5.1%	-9.2%	9.0%	1.6%	-11.0%	-8.3%	
											< 10% Green
1 = Agent on duty during the hour							**Average Difference from Target:**				**13.2%**
											> 15% Red

Besides making updates to the DSS based on feedback from the call center manager, other features have been contemplated. In order to maintain performance proactively, a demand forecasting model will be considered. Because each agent is expected to differ in average service times, a tracking model will be considered that automatically updates average service times for new agents whose service times will change over time. Performance outcomes may also be tracked using, for example, a control chart for outcomes associated with caller waiting times. Finally, although flexibility is important to the manager, an optimization routine will be considered for the DSS.

REFERENCES

Aksin, Z., Armony, M., & Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management, 16*(6), 665-688.
Akşin, O., & Harker, P.T. (2003). Capacity sizing in the presence of a common shared resource: Dimensioning an inbound call center. *European Journal of Operational Research, 147*(3), 464-483.
Aktekin, T., & Ekin, T. (2016). Stochastic call center staffing with uncertain arrival, service and abandonment rates: A Bayesian perspective. *Naval Research Logistics), 63*(6), 460-478.
Avramidis, A.N., & L'Ecuyer, P. (2005). Modeling and simulation of call centers," *Proceedings of the 2005 Winter Simulation Conference*, 144-152.
Barton, R.R. (2015). Tutorial: simulation metamodeling. *Proceedings of the 2015 Winter Simulation Conference*, 1765–1779.
Bielen, F., & Demoulin, N. (2007). Waiting time influence on the satisfaction-loyalty relationship in services. *Managing Service Quality: An International Journal, 17*(2), 174-193.
Bouzada, M.A.C. (2009). Dimensioning a call center: Simulation or queue theory? *Journal of Operations and Supply Chain Management 2*(2), 34-46.
Cezik, M.T., & L'Ecuyer, P. (2008). Staffing multiskill call centers via linear programming and simulation. *Management Science, 54*(2), 310-323.
Corlu, C.G., Maleyeff, J., Yang, C., Ma, T., & Shen, Y. (2021). Decision support system with simulation-based optimization for healthcare capacity planning. *Proceedings of the 10th Simulation Workshop of the Operational Research Society*, Leicestershire, UK.
Chromy, E., Diezka, J., & Kavacky, M. (2012). Markov model M/M/M/K in contact center. *Advances in Electrical and Electronic Engineering, 10*(4), 264-269.
Davis, M.M., & Vollmann, T.E. (1990). A framework for relating waiting time and customer satisfaction in a service operation. *The Journal of Services Marketing, 4*(1), 61.
Doomun, R., & Jungum, N.V. (2008). Business process modelling, simulation and reengineering: Call centres. *Business Process Management Journal, 14*(6), 838-848.
Dudin, S., Kim, C., Dudina, O., & Baek, J. (2013). Queueing system with heterogeneous customers as a model of a call center with a call-back for lost customers. *Mathematical Problems in Engineering, 2013*, 1-13.
Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management, 5*(2), 79-141.
Gans, N., Shen, H., Zhou, Y., Korolev, N., McCord, A., & Ristock, H. (2015). Parametric forecasting and Stochastic programming models for call-center workforce scheduling. *Manufacturing & Service Operations Management, 17*(4), 571-588.
Garcia, D., Archer, T., Moradi, S., & Ghiabi, B. (2012). Waiting in vain: Managing time and customer satisfaction at call centers. *Psychology, 3*(2), 213-216.
Gong, J., Yu, M., Tang, J., & Li, M. (2015). Staffing to maximize profit for call centers with impatient and repeat-calling customers. *Mathematical Problems in Engineering, 2015*, 1-10.
Grobbelaar, R.L., Roodt, G., & Venter, A. (2004). Critical competencies to promote a customer service core capability in a human resources call centre. *SA Journal of Human Resource Management, 2*(3), 16-24.
Harrison, J.M., & Zeevi, A. (2005). A method for staffing large call centers based on stochastic fluid models. *Manufacturing & Service Operations Management, 7*(1), 20-36.

Hasija, S., Pinker, E., & Shumsky, R.A. (2010). OM practice - work expands to fill the time available: Capacity estimation and staffing under Parkinson's law. *Manufacturing & Service Operations Management, 12*(1), 1-18.
IBISWorld (2020). Telemarketing & call centers in the US market size 2002–2025. New York: IBISWorld.
Kanavetas, O., & Balcıoğlu, B. (2018). The “Sensitive” Markovian queueing system and its application for a call center problem. *Annals of Operations Research*, 1-14.
Kim, C., Dudin, S., Taramin, O., & Baek, J. (2013). Queueing system MAP/PH/N/N + R with impatient heterogeneous customers as a model of call center. *Applied Mathematical Modelling, 37*(3), 958-976.
Kim, C., Klimenok, V., & Dudin, A. (2016). Priority tandem queueing system with retrials and reservation of channels as a model of call center. *Computers & Industrial Engineering, 96*, 61-71.
Labach, E. (2010). Improving customer retention through service quality at call centers. *International Journal of Management and Information Systems, 14*(3), 71-76.
Liao, S., Koole, G., Van Delft, C., & Jouini, O. (2011). Staffing a call center with uncertain non-stationary arrival rate and flexibility. *OR Spectrum, 34*(3), 691-721.
Ma, J., Kim, N., & Rothrock, L. (2011). Performance assessment in an interactive call center workforce simulation. *Simulation Modelling Practice and Theory, 19*(1), 227-238.
Mehrotra, V., & Fama, J. (2003). Call center simulation modeling: Methods, challenges, and opportunities," *Proceedings of the 2003 Winter Simulation Conference*, 135-143.
Millhiser, W.P. & Veral, E.A. (2019). A decision support system for real-time scheduling of multiple patient classes in outpatient services. *Healthcare Management Science, 22*(1), 180-195.
Munoz, D.A., & Bastian, N.D. (2016). Estimating cross-training call center capacity through simulation. *Journal of Systems Science and Systems Engineering, 25*(4), 448-468.
Örmeci, L. & Akşin, Z. (2010). Revenue management through dynamic cross selling in call centers. *Production and Operations Management, 19*(6), 742-756.
Peevers, G., McInnes, F., Morton, H., Matthews, A., & Jack, M. (2009). The mediating effects of brand music and waiting time updates on customers' satisfaction with a telephone service when put on-hold. *The International Journal of Bank Marketing, 27*(3), 202-217.
Salvador, S., & Chan, P. (2004). Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, 576-584.
Sariyer, G. (2018). Sizing capacity levels in emergency medical services dispatch centers: Using the newsvendor approach. *The American Journal of Emergency Medicine, 36*(5), 804-815.

DECISION SCIENCES INSTITUTE

Best Matching Protocol-based Collaborative Crowdsourcing Contracting Evaluation

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Fangyu Liu
University of North Texas
Email: ys_liu999@163.com

Nagi Gebraeel
Georgia Institute of Technology
Email: nagi@gatechl.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

Product fulfillment through crowdsourcing has been an emerging trend towards Industry 4.0. This paper envisions collaborative crowdsourcing contracting decisions among design and manufacturing agents as a best-matching problem. In the product fulfillment process, it requires agents and organizations to dynamically form and configure a supply chain with design and manufacturing capabilities. To identify preferrable agents to customers, this paper develops a methodology for agent selection in a collaborative network by developing a best matching protocol (BMP). Agent selection involves criteria evaluation on both engineering and operational performance, and the proposed BMP focuses on operational performance evaluation through decision tree learning.

KEYWORDS: Open design, Open manufacturing, Crowdsourcing, Collaborative negotiation, Supply contracting, Mass customization.

INTRODUCTION

Manufacturing industries of the future are poised for more disruptive changes that are impacted by sustainability issues, consumer attitudes and behaviors, digitization, 3D printing, emerging automation and technology, etc. Industry 4.0 has been advocated as the Fourth Industrial Revolution that indicates the ongoing automation of traditional manufacturing and industrial practices, using modern smart technologies, large-scale machine-to-machine communication and the Internet of Things (Berger, 2014). The Industry 4.0 manifesto was first published by the German National Academy of Science and Engineering (Kagermann, 2013), suggesting that the "Industry 4.0 is a new level of value chain organization and management across the lifecycle of products". There are four basic pillars of Industry 4.0, including smart solutions, smart innovations, smart supply chain and smart factory (Capgemini, 2014). Smart solutions involve smart products and smart services, which coincides with a significant trend of transforming and expanding the manufacturing sector towards mass customization and open manufacturing as a service.

(1) Mass customization and open enterprise: At the frontend of manufacturing, firms are confronted with challenges for satisfying individual customer needs while managing product variety in product development more efficiently than competitors (Brettel et al., 2014). The extent of market-of-one has been foreseen as a potential driving force for the next transformation of the global economy (Pine, 2009), leading to paradigm shift from traditional mass production to mass customization and personalization (Tseng et al. 2010). Large variety in the market and product domain propagating to the manufacturing domain implies higher variations of production processes, causing complex technology portfolio (ElMaraghy et al., 2013) and restorative supply chain operations in a circular economy (Kirchherr et al., 2017). As a result, future manufacturing of Industry 4.0 needs to utilize information sharing to cooperate with a peer of suppliers through a series of outsourced fabrications to finish final products within the lead time (Sahin and Robinson, 2005). Instead of owning excess production capacities to hedge against demand fluctuation, companies tend to adopt an open business model (Chesbrough and Appleyard, 2007) by implementing manufacturing capabilities and resources as a scalable and changeable production network to achieve quick response and adjust capacities (Freitag et al., 2015) through an agile enterprise structure with extensive capability arsenal (Kortmann and Piller, 2016). The open business model allows the creation of new complementary links throughout a value chain, which explicitly arranges the stakeholders along with positions of value creation, delivery and capturing (ElMaraghy and ElMaraghy, 2014). This openness in manufacturing enables mass cooperation of the manufacturers to construct a highly interactive manufacturing network that relies on the cooperative collaboration mechanism, which eventually empowers easy installation of new technologies and scaling up of manufacturing capabilities.

(2) Crowdsourcing product fulfillment: Crowdsourcing has emerged as a promising means of open business operations by transcending organizational boundaries to leverage resources and capabilities across distributed stakeholders (Kohler, 2015). Different from the conventional strategy of outsourcing in supply chain management that emphasizes how to delegate a task to a designated agent, crowdsourcing utilizes an open call to a crowd for maximally exploiting the external resources (Bücheler and Sieg, 2011). Among many perspectives of crowdsourcing, Industry 4.0 emphasizes a cyber-platform-driven perspective to peel out the coordinating and negotiating responsibilities from the crowdsourcing agents while embedding an open manufacturing business model as the foundation of product innovation and development. The crowdsourcing cyber platform facilitates a manufacturer to explore external knowledge and resource by coordinating the activities of design and manufacturing among the stakeholders through a collaborative manufacturing fulfillment network.

The impacts of design and manufacturing through crowdsourced on potential industrial and societal benefits are profound. The cutting-edge information and communications technologies and industry trends empower manufacturing to be fulfilled over a crowdsourcing cyber platform to transcend the partners' borders extensively and build up information exchanging networks. Integration of smart sensors and the networked manufacturing systems equips firms with a cyber-physical-human environment, where a synergy of Internet of Things, big data analysis, and machine intelligence with legacy manufacturing technologies such as computer integrated manufacturing systems, supply chain management and production logistics have stimulated a gigantic manufacturing technology advancement in future manufacturing. The ultimate goal is to achieve competitiveness in collaboration across multiple entities towards an enterprise with an open yet virtual architecture (Crawford, 2018), which is in line with the vision of Industry 4.0 characterized as decentralization and digitalization (Schwab, 2017; Schuh et al., 2014).

(3) Design and manufacturing as a service: Future product fulfillment in Industry 4.0 is equipped with ubiquitous connectivity in the manufacturing environment, allowing collection of significant volumes of dispersed information to support distributed decision making in fulfilling manufacturing tasks (Monostori et al., 2016). The new open design and manufacturing capabilities enabled by crowdsourcing platform will create opportunities for transforming and expanding the manufacturing sector by developing intelligent cognitive assistants to perform as decision support systems to facilitate the fulfillment of manufacturing as a service (Li et al., 2018b). The compelling need for accommodating a dynamic and collaborative network of manufacturing services has a broader implication for a service-oriented paradigm to be deployed as X-as-a-service and extended to the entire design and manufacturing regime to act as service manufacturing (Kusiak, 2019). Another implication is dedicated to social manufacturing (Jiang et al., 2016a) that aims to take advantage of the interactive relationships among the manufacturer crowds to foster a manufacturing service network as an autonomous organizing process. The trend of cloud-based design and manufacturing offers a framework of connecting smart entities across a population of companies, thus enabling a demand-capacity matching mechanism to serve collaborative product realization (Schaefer, 2014). Crowdsourcing enables product fulfillment as a service through a cyber-platform-based design and manufacturing system, which is organized as a cloud-based resource sharing mechanism among the manufacturer crowds while engaging more manufacturer population in the smart service network (Kaihara et al., 2017).

The emerging cyber-physical-human production systems will provoke changes in many ways for future design and manufacturing concerning product fulfillment as a service in the factory of the future. Leading experts expect less basic, repetitive work but more ambitious tasks in collaboration with the crowdsourcing platform, and thus the factory of the future will not be deserted but organized as a network of crowdsourcing platform-driven design and manufacturing services (Mattsson et al., 2012). It can be envisioned that the central theme of crowdsourcing product fulfillment lies in the new design and manufacturing capabilities empowered by the crowdsourcing cyber platform that performs as intelligent cognitive assistants to engage a large population of designer and manufacturer crowds in an extended smart service network.

(4) Collaborative-negotiation in crowdsourcing contracting: Successful application of open design and manufacturing requires collaboration among external partners, for which product fulfillment flow management is of primary importance. The open innovators, open designers, and open manufacturers are all engaged through an inter-organizational network and their crowdsourcing relationships are contractually tied to collaboration for fulfilling different knowledge and capabilities along with a coherent product fulfillment flow (Simard and West, 2006). Such a crowdsourcing contracting mechanism is akin to traditional supply contracting that formally formulates the transactions between the stakeholders to pursue the coordination of diverse decision makers and organize them into supply chain networks (Giannoccaro and Pontrandolfo, 2004).

There is a stream of research of negotiation system for coordination of distributed enterprises, which are coinciding with the product fulfillment process (Mansouri et al., 2012). This proposing system entails a bilateral negotiation scheme coincides with a supply contract with an emphasis on the design of the efficient negotiation mechanisms, protocols, and strategies (Shin and Jung, 2004). In practice, every organization and entities in the supply chain networks are operating in heterogeneous environments with different objectives and constraints (Swaminathan, 1996).

Since it is observed that a successful crowdsourcing decision-making process requires diversity and independence of the individuals in the crowds (Surowiecki, 2004), the crowdsourcing contracting is more challenging than conventional supply contracting.

Besides, any entities involved in the product fulfillment scheme must be considered for their cohort behavior, instead of only their individual operations, in order to achieve the general functionality along the product fulfillment flow and to negotiate with their peers to find compromised solutions (Sadeh et al., 2001). This indicates that the essential issue of crowdsourcing contracting lies in collaborative negotiation along the decision-making flow of product fulfillment (Jiao et al., 2006).

The collaborative negotiation process is generally divided into three consecutive phases, namely inviting, bidding and awarding. Corresponds to a crowdsourcing product fulfillment scheme, the inviting is the crowdsourcer sends an open call to a crowd for solutions or capabilities of an independent crowdsourcing subtask. Subsequently, the crowds will solve the subtasks with a bid. The collection of the bids from the crowds can be described as a contest to the rewards, under a scheme of highest-bid-wins, considering the performance or efforts to the original subtask. A designated collaborative negotiation contracting scheme should serve not only the interaction among crowdsourcing entities but also the motivation of the crowds and the quality of the final products. Such requirement implies an effective contracting evaluation mechanism to assess the bids from crowds' contest.

Towards this end, this paper is geared towards a contracting mechanism to support the crowdsourcing product fulfillment process. In line with best practices of order fulfillment decisions in supply networks, the paper propose to formulate crowdsourcing contracting as a best-matching problem (Xu et al., 2009). The most important activity of the crowdsourcing product fulfillment process is the selection of the fulfilling agents in an open innovation, open design and open manufacturing environment. The selection of the agent involves multiple criteria based on quantitative and qualitative metrics and requires several self-interested agents and organizations to dynamically form and configure a supply chain with appropriate design and manufacturing capabilities. To support selection of a crowdsourcing agent in a collaborative, geographically distributed network, a best matching protocol (BMP) is developed to evaluate agent operational performance regarding implicit criteria based on decision tree learning for monotonic classification, since engineering performance evaluation is a well-explored field.

LITERATURE REVIEW

Open Business Model and Crowdsourcing

Open business model is defined as utilizing the external partners' assets to develop own business model (Chesbrough, 2006a). Open business model enhances the firm's efficiency by leveraging external resources in value-creating processes and achieving high utilization of not only the firm's key assets but also the external partners' resources in value capture process (Chesbrough, 2007). As a later supplementary of the open business model, open innovation is applied to depict the distributive innovation process based on purposively managed flows across the organization's boundaries (Bogers et al., 2017). Open innovation has been recognized as an opposite of the traditional vertical integration model, regarding develops and distributes the products by one firm (Chesbrough et al., 2006). Open innovation horizontally structures a dynamic interaction network of various clusters of autonomous firms throughout the product innovation process (Dhanaraj and Parkhe, 2006). Moreover, from a platform-based view, an increasing amount of the industry examples organize the firms as a central platform structure, the core firm seeks the inflow of the external knowledge, while the surrounding firms outflow their knowledge (Gawer and Cusumano, 2014).

Open design and manufacturing are other instantiations of the open business model, which is introduced to depict the collaboration with the external designers and manufacturers crowds through the product design and manufacturing processes, respectively (Bauwens, 2009). The conception of open design origins from the open source method from the software industry, which has created the legends includes Linux and Wikipedia (Weber, 2004). Open design often entails the collaboration of external designers to design the subsystem, which can be integrated by an open architecture harmoniously (Vallance et al., 2001). The practice of open design is enabled by a designer community and the internet-based communication technologies, which significantly reduce the cost of virtually team structuring and collaboratively operating (Koch and Tumer, 2009). Manufacturing plays an indispensable role in the physical products' value realization process (Koufteros et al., 2014). The implementation of the open business model through manufacturing is identified as the critical challenge on opening fulfillment processes of physical products (Maurer and Scotchmer, 2006). However, the increasing digitalization and data-centricity enable the open business model to propagate from software development to the fulfillment of tangible products (Raasch et al., 2009).

Recent studies highlight open manufacturing and cloud manufacturing as critical perspectives to collaborate with a crowd of manufacturers. Open manufacturing entails an enterprise structure which integrates the knowledge of manufacturing from a distributed manufacturer community (Maurer and Scotchmer, 2006). Cloud manufacturing integrates the network, cloud computing and smart manufacturing technologies into the transformation process of manufacturing resources and capabilities in the manufacturing services (Zhang et al., 2014). Both approaches ease the collaboration with manufacturer crowds in the context of the physical product fulfillment. It is observed that with the support of an open-source platform, the crowds of designers and manufacturers can be configured as a collaborative team to fulfill the physical product (Banerjee et al., 2015). This trend implies a paradigm shift to relying on the knowledge and resources from the crowds to fulfill the product development.

Among the accesses to the external knowledge and resources, crowdsourcing is highlighted and described as the open innovator broadcasts the problem to the crowd and select the best solution, instead of outsourcing a problem to a designated agent or solving it internally (Afuah and Tucci, 2012; Howe, 2006). Thanks to the wisdom of crowds, the collection of intelligence from the crowds show equal to or superior to the few elite (Leimesiter et al., 2009). From the perspective of participants, several participating motivations has been explored, includes self-market or promotion, tangible or intangible compensation, social fames and reputation, to name but a few (Bayus, 2010).

Since the crowdsourcing mechanism is confirmed to benefit the technical problem solving, several researches gear forward the formulation of the crowdsourcing. Surowiecki (2004) identifies four prerequisites to ensure the successful crowdsourcing decision-making: 1) diversity, each participant can offer unique knowledge or capabilities; 2) independence, to avoid the influence from peers; 3) decentralization, the information is sharable to the participants to locally process; 4) aggregation, the fulfilled crowdsourcing task can be aggregated to a collective delivery. Bonabeau (2009) divides the crowdsourcing decision-making processes into two stages, generating a set of possible solutions and evaluating each solution. The crowd generates solutions independently and decentrally to ensure the cognitive diversity, then aggregates the set of best solutions (Rosen, 2011).

However, with an increasing interest in these fields, there is limited research shed light on the elaboration of the crowdsourcing product fulfillment through an open business model. Several essential difficulties in this field are highlighted, which includes partner’s participation payoffs, value capture ability, and the prosperity of the collaborator crowds, trust development scheme, and coordination levels. (Appleyard and Chesbrough, 2017; Brunswicker and Chesbrough, 2018; Van de Vrande et al., 2009; Saxton et al., 2013). Furthermore, there is still an absence of formulated open design and manufacturing workflow which accommodates the crowdsourcing mechanism (Bogers, Zobel, 2017).

Enablers of Crowdsourcing Product Fulfillment

The current achievement of the information and communications technologies (ICT) impulse the emergence of crowdsourcing. The way of a crowdsourcer reaching crowds to allocate a product fulfillment task relies on the exponential growing ICT. Many disruptive technologies have been advocated, and continuously emerging, for advancing manufacturing and service operations, such as cloud computing, industrial Internet of Things, cyber-physical systems (CPS), to name but a few. These new technologies are nowadays penetrating many manufacturing and service sectors as critical enablers for the companies to innovate their businesses and product fulfillment for better achievement of customization and quick response. For example, Industry 4.0 envisions a synergy of IoT and CPS in a manufacturing environment to make manufacturing systems smarter and more autonomous, leading to high agility and flexibility of the production processes (Weyer et al., 2015). Such technological trends bring pervasive connectivity to the manufacturing environment, allowing collection of significant volumes of real-time information (Monostori et al., 2016).

Owing to advantages in collaboration across multiple entities towards the competitiveness of the companies, decentralization has been recognized as one core characteristic of Industry 4.0 (Schuh et al., 2014). This trend emphasizes the coordination which highly depends on the service and communication platforms throughout the product fulfillment process (Richardson, 2016). The emerging cloud computing facilitates the real-time collaboration of various stakeholders to form intelligent networks for efficient product fulfillment (Xu, 2012). Moreover, smart manufacturing technologies like additive manufacturing significantly reshape the product realization process with digitalized knowledge (Ratto and Ree, 2012) and networked supply chains (Holland et al., 2017). Thanks to advanced CAX software, collaborative product fulfillment opens the access of a design team’s simultaneously modification of the digital files with the cohesive organization of distributed knowledge to ensure successful outcomes (Zhen et al., 2011). Additionally, by applying cloud-computing technologies, the cloud-based design and manufacturing (CBDM) builds up the bridges between the individual design and manufacturing partners while offers the opportunities of inter-organizational real-time communication to synchronize product fulfillment activities (Wu et al., 2013a). Because of the support from the cyber platform, the paradigm of CBDM enables the rapid finding of the optimal resource allocation among the crowds for various demands and smooths the collaboration path (Wu et al., 2013b). Meanwhile, the CBDM helps the design team exploiting a set of various and distributed available manufacturing resources for efficiency enhancement of the realization of product design (Wu et al., 2015).

In addition to the connection to the decentralized designer and manufacturer crowds, a platform to aggregate the decentralized task to the delivered product is critical to the crowdsourcing product fulfillment. Product family design and platform-based product development method offers a systematical method to manage the variety of the delivered products and subtasks (Jiao et al., 2007). Product platform includes three aspects: 1) the modular architecture; 2) the interfaces; and

3) the standard which the modules must follow (Baldwin and Clark, 2000). The modularity offers the independence of the subtasks, which benefit the fulfillment of customer needs and the success of the crowdsourcing mechanism (Jiao and Tseng, 2004). The interfaces and standard ensure the collection of the subtask and the efficiency of product development (Meyer and Lehnerd, 1997).

The trends of digitization of product design process and smart manufacturing technologies enable the implementation of product platforms in the crowdsourcing process (Boisseau et al., 2018). In the meantime, big data analytics has been introduced to the industry to support decision-making in an information explosion era (Brown et al., 2011). Application of big data can not only empower intelligent decision making in product design and production but also ushers in the socialization of product design with predictive analytics of the suppliers' performance (Li et al., 2015). The concept of CPS is generated to interconnect the physical manufacturing equipment and the data flow, for the purpose of achieving an agile and intelligent manufacturing system (Lee et al., 2015). This interconnectivity can also boost the replacement of the central re-planning process in the production system, with an autonomous process of the reconfiguration of product and production units (Rosen et al., 2015). The synergy of these cutting-edge technologies consists of a set of enablers of crowdsourcing product fulfillment.

PRODUCT FULFILLMENT THROUGH COLLABORATIVE CROWDSOURCING

Workflow of Collaborative-Crowdsourcing Product Fulfillment

The collaborative-crowdsourcing product fulfillment provides an integrated path of external partners into all the activities in value creation and value capture, such as product design and manufacturing. Following the model of tournament-based crowdsourcing, the fulfillment of a product is achieved by decomposing the original task to a set of subtasks and holding a corresponding series of tournaments to collect the fulfillment effort. This process results from the collaboration of multi-parties in four physical domains: open product innovation domain, collaborative-crowdsourcing product fulfillment platform, open design domain, and open manufacturing domain. Inspired by the axiomatic design model, the workflow of collaborative-crowdsourcing product fulfillment through open design and manufacturing is illustrated below.

The open innovator falls into the open product innovation domain. As initiating stakeholders, they build up the connection with customers and collects CNs. They subsequently specify FRs with engineering consideration. The product fulfillment process relies on the collaboration of design agents and manufacturing agents, which can be accessed by a crowdsourcing platform via a tournament-based crowdsourcing. The crowdsourcing platform has design and manufacturing platform agents, which decompose the product fulfillment request to a series of crowdsourcing tasks. These tasks are sent to the design and manufacturing agent crowd in the form of "Request for Quotation" as an open call. According to their capability constraints and operation consideration, some of them participate in the tournament by proposing bids for the task. These bids can be evaluated and selected to form a set of best solutions, which can be aggregated into a final product fulfillment solution.

Design agents then outsource the design solutions to fulfill FRs. Stakeholders in the open manufacturing domain then plan and execute the manufacturing plans with the consideration of process capability and resource utilization limitation. The collaborative-crowdsourcing product fulfillment platform is the fourth domain. It is a bridge to the open product innovation domain on the front end and the open design and manufacturing domain on the back end. It has two virtual fields, including crowdsourcing information management and crowdsourcing supply contracting mechanism. The crowdsourcing information management field is the interface to the open product innovation domain. The crowdsourcing supply contracting mechanism is responsible for negotiating and contracting with design agents and manufacturing agents. Correspondingly, there are two stakeholder clusters: design platform contracting brokers and manufacturing platform contracting brokers. They host the tournament to search for the best solutions for the design subtasks and manufacturing subtasks.

The CNs can be collected and saved as a set of the customer orders, including the expectations of final delivery products. These orders are processed in open product innovation domain, where the initiating stakeholder develops FRs with engineering concerns and related technologies. FRs are formulated and saved as product specs and structured as a Cartesian product, which depicts the combination of specific functional requirements and interrelationships among them. In order to host the design tournament, design platform agents decompose the Cartesian product to the design request for quotation (RFQ). Each design subtask includes a set of design subtasks to fulfill a portion of FRs. Based on the decomposition, the bidding design agents receive product specs from the open call, analyze the corresponding tasks, and response with the design bid. Bids from all design agents in this cluster are collected in the bid set. Design platform agents collect design bids in the design bid set and evaluate them based on the corresponding product specs. After bid evaluation, they select the preferred bids as the winner of the tournament for each RFQ and send the result back to the crowdsourcing information management field in the form of the design spec set, which contains all the selected design bids in every design agent cluster. Similarly, based on the understanding of the manufacturing industry, design specs are also structured as the product structure, which depicts the collection of design requirements and the configuration of a product. The following activities are similar to the design bidding process.

Different from the "cascading" model, the product fulfillment process in open enterprise is in the form of "zigzagging". The reason for the difference is the involvement of the external partners, and the fulfillment is achieved by the collaboration of all stakeholders in the fulfillment process. However, this kind of collaboration is forged in the form of contracting, and the negotiation involves the supply contracting which coordinates the product design and the material flow (Jiao et al., 2006; Subramanian et al., 2009). Thus, the product fulfillment process in open enterprise can be characterized as collaborative negotiation-based product fulfillment process.

Crowdsourcing Contracting Process

The tournament in the collaborative-crowdsourcing product fulfillment is achieved by crowdsourcing contracting, which exists in three stages of the fulfillment processes. Firstly, design platform agents negotiate with design agents to search for the best design solution. Secondly, manufacturing platform agents negotiate with manufacturing agents to allocate manufacturing tasks as a synchronized fulfillment network with the best performance. The design and manufacturing supply contracting mechanisms are similar, and a generalized contracting process for design and manufacturing is illustrated below.

The contracting between crowdsourcing platform and crowdsourcing crowds can be decomposed into three inter-domain activities: 1) RFQ broadcasting; 2) bid proposing; 3) contracting. Based on the role in the contracting mechanism, platform contracting agents can be categorized into project configuration manager and negotiation brokers. Negotiation brokers can be further decomposed into the agent invitation broker and the bids evaluation broker. Based on the status of applied policies, crowdsourcing agents can be categorized into bidding design cluster and non-bidding design agents.

The platform receives the requirement structure and delivers the product spec sets as the result of the negotiation. The collaborative relationship among is formed by the crowdsourcing supply contracts. The project configuration manager receives product fulfillment requirements, decomposes and restructures them as several RFQs. After that, these RFQs will be sent to negotiation brokers. The agent invitation broker takes in charge of the invitation to issue the RFQ and collect bids. According to the one-to-one relationship of RFQ and crowdsourcing agent cluster, each broker takes in charge of the invitation a cluster of bidding design agents for a specific RFQ.

Every bidding agent proposes a bid. The bids are collected in a bid set. All bids are sent to one-to-one to the bid evaluation broker. After the evaluation process, the preferred crowdsourcing agent is selected from every agent cluster, and the contracts are generated subsequently. The crowdsourcing supply contracts will be sent to open design domain and open manufacturing domain. The contracts are also copied to the project configuration manager, who synthesizes bids as the product spec set. These product fulfillment solutions are sent to crowdsourcing information management field and saved.

Contracting Evaluation

The evaluation brokers serve as the search engine of the crowdsourcing system. The results of the evaluation influence the quality of final products as well as the motivation of the open designers and manufacturers. The challenges for evaluating a bid lie in three aspects. Firstly, the crowdsourcing contracting evaluation is characterized as a large-scale multi-criteria decision-making problem. Different from the traditional outsourcing which invites designated partners to participate in the product fulfillment process, crowdsourcing relies on the wisdom of crowds, which implies the crowd can generate a large volume of solutions (Lakhani and Panetta, 2007). Meanwhile, the trade-offs of the crowdsourcing tasks are reflected by various of conflicting criteria in evaluation. After single-criterion evaluation, the result should be ready to be aggregated for comparison. To sum up, the evaluation is a complex problem, a generic and formulated evaluation scheme is essential to handle the scale of evaluation.

Secondly, a stream of uncertainty is inevitable along the crowdsourcing product fulfillment. From a design perspective, this uncertainty can be traced from the subjectiveness lying in the evaluation process and the variation of system performance (Jiao and Tseng, 1998; Siskos et al., 1984). In practice, the experts conduct evaluation based on their heuristic “rule of thumb”, which has been historically done on an ad hoc basis (Thurston and Crawford, 1994). Establishing a model of the preference of the bids and the decision-making in the evaluation process to serve the contracting mechanism is critical to the realization of collaborative-crowdsourcing product fulfillment. From the manufacturing perspective, the performance of the production system shows strong dynamic and stochastic characteristic in the real manufacturing environment. Such characteristics are shown in the fluctuation of the throughput time, tolerance and rejection rate. In addition, the evaluation of the contracting is in the early stage, which implies the design and manufacturing solutions are subjected to slight changes in later process. A method to mimic the uncertainty of the performance is critical in the development of evaluation mechanism.

Thirdly, the evaluation is observed as a double-folded process since crowdsourcing aims to fulfill requests. It is observed that the crowds in the crowdsourcing activities show a return of the vast number of noises (Keen, 2007). An evaluation mechanism should ensure the performance of the delivered solution which can target on the requirements. From the design perspective, the evaluation should pursue the minimized deviation of system performance to design requirements to obtain the maximized customer satisfaction. Similarly, the manufacturing evaluation should minimize the deviation of the production performance to the supplying fulfillment requirements to guarantee the performance of the reconfigured system.

DECISION TREE LEARNING FOR OPERATIONAL PERFORMANCE EVALUATION WITH HISTORY MONOTONE ORDINAL MEASUREMENT

Operational Performance Evaluation with Intangible Criteria

Based on the characteristics of the criteria, evaluation problems can be divided into both tangible criteria evaluation and intangible criteria evaluation. The former indicates approaches to quantify the object performance according to the given criteria, like most of the evaluation for engineering performance. However, not all performance can be quantified with numeric values based on the criteria requirements. In this scenario, imprecise linguistic words can be used to fuzzily evaluate the performance, like in some criteria for business or operational performance.

Because the objective of evaluation is to compare the performance of different objects from a certain perspective, to replace absolute numeric measurement with ordinal measurement for describing the degree of satisfaction is an approach to intangible criteria evaluation. By doing so, objects can be ranked and ordered based on their performance. This will not violate the inexplicitness or fuzziness regarding the intangibility of a criterion. Therefore, as long as the contracting brokers keep the evaluation mechanism consistent, the ranking or ordering of the object performance will be based on the same logic, making the objects at different time points comparable. Furthermore, such evaluation is supposed to be monotone, which means when one object dominates another, the evaluation of the dominator will not be worse than the dominated one.

In this regard, the evaluation of intangible criteria can be modeled as a monotonic classification problem. To address this problem, the rank entropy-based decision tree is used to learn the evaluation mechanism under intangible criteria, in which rank entropy is used to branch the decision tree based on the data monotonicity.

Definition of Monotonic Classification

Monotonic classification refers to ordinal classification problems with the monotonic constraint. Let A be an instance space where p is the number of attributes, which can be noted as $A = A_1 \times A_2 \times \ldots \times A_p$. Let $U = \{x_1, x_2, \ldots, x_n\}$ be a set of objects in the instance space A, with D being the ordinal decisions or labels of these objects. The value of the attribute or the decision related to x_i can be expressed as $v(x_i, a)$ or $v(x_i, D)$, where $a \in A$. In the ordinal relation, $\leq$ is used to describe no worse than between two objects. For example, x_i is no worse than x_j in terms of D can be noted as $v\left(x_j, D\right) \leq v(x_i, D)$ or $x_j \leq_D x_i$. Usually, x_i dominates x_j refers to that every attribute value of x_i is no worse than x_j. Based on this concept, a predicting function f that relates A to D can be expressed as below:

$$f: U \rightarrow D \tag{1}$$

The monotonic constraint is then defined as below, which should always be satisfied in monotonic classification:

$$x_i \leq x_j \Rightarrow f(x_i) \leq f(x_j), \forall x_i, x_j \in U \tag{2}$$

In other words, if x_j is dominated by x_i, the decision of x_j will not be worse than that of x_i, not vice versa.

Rank Entropy-based Decision Tree

Rule extraction from monotonic data attracts some attention from the domains of machine learning and decision analysis. Decision tree induction is an efficient, effective, and understandable technique for rule learning and classification modeling (Quinlan, 1993), where a function is required for evaluating and selecting features to partition samples into finer subsets in each node. The rank entropy measure originates from Shannon's information entropy, which is robust in evaluating features of the monotone dataset (Hu et al., 2011). Also, this measure reflects the ordinal structures in monotonic classification. Therefore, a decision tree algorithm based on the rank entropy measure is used in this study.

Some preliminary definitions are given for introducing the rank entropy-based decision tree.

$$[x_i]_B^{\leq} = \{x_j \in U | x_i \leq_B x_j\}, \text{ where } B \subseteq A \tag{3}$$

$$[x_i]_D^{\leq} = \{x_j \in U | x_i \leq_D x_j\} \tag{4}$$

Equation (3) and (4) describe objects no worse than x_i in terms of attributes or decision. Similar to the concept of the information entropy, the ascending rank entropy of an object set U is defined as below:

$$RH_B^{\leq}(U) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|[x_i]_B^{\leq}|}{n} \tag{5}$$

The ascending rank joint entropy of an object set U is defined as

$$RH_{B \cup D}^{\leq}(U) = -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|[x_i]_B^{\leq} \cap [x_i]_D^{\leq}|}{n} \quad (6$$

The ascending rank mutual information (RMI) of an object set U is defined as

$$RMI^{\leq}(B, D) == -\frac{1}{n}\sum_{i=1}^{n} \log \frac{|[x_i]_B^{\leq}| * |[x_i]_D^{\leq}|}{n * |[x_i]_B^{\leq} \cap [x_i]_D^{\leq}|} \tag{7}$$

The rank entropy-based decision tree uses the RMI value for the branch operation. Essentially, RMI describes the degree of monotonicity between the attribute set B and the decision set D.

The workflow of the rank entropy-based ordinal decision tree can be referred to (Hu et al., 2011).

A Worked Example of Decision Tree Learning for Agent Service Evaluation

To evaluate part of implicit criteria, monotonic classification is conducted with the rank entropy-based decision tree. In this study, an example of customer service evaluation is done as one group of criteria regarding the business performance of the crowdsourcing agents. The defined customer service has 4 rate levels: A, B, C, and D, where A represents the top-level customer service and D is the worst. The rating depends on the evaluation value of four criteria: the support responsiveness, how easy is the problem solved, the service attitude, and the satisfaction of this service. Each of these criteria is rated with integers $\{0,1,2,3\}$, where a higher value represents more positive feedback. Based on the above formulation, a monotone ordinal dataset is constructed, as shwon in Table 1. The dataset is created with an algorithm that generates unstructured monotone ordinal data (Potharst et al., 2009). The data consists of 60 entries, and A, B, C, D have $16, 15, 15, 14$ entries each.

Table 1: The Monotone Dataset for Customer Service Evaluation

A	0332, 1213, 2133, 2222, 2233, 2330, 2333, 3203, 3213, 3231, 3232, 3311, 3313, 3321, 3322, 3323
B	0213, 0230, 0320, 1132, 1303, 1311, 2122, 2131, 2203,2221, 2311, 3122, 3200, 3212, 3221
C	0111,0202,0212,1102,1121,1131,1211,2013, 2023,2103,2111,2120,2130,3033,3121
D	0013,0020,0022,0120,0200,0201,1003, 1011,1210,2011,2020,3001,3012,3100

The result of the rank entropy-based decision tree from the above dataset is a decision tree with 31 leaves. The generated 31 leaves correspond to 31 rules for customer service classification.

Managerial Implications

Following the method of tournament-based crowdsourcing, this paper shows the process of reaching external partner crowds via a platform and instantiates the tournament to search the best-performed solution. In addition, the proposing evaluation mechanism provides an approach to constructing the collaborative-crowdsourcing team. Meanwhile, the collaborative-crowdsourcing product fulfillment offers a bridge to link crowdsourcing innovators, crowdsourcing designers, and crowdsourcing manufacturers to the customers. They share their core competitiveness and capabilities complimentarily. Therefore, they have higher possibility to achieve economies of scale and ease the application of the emerging technologies.

CONCLUSIONS AND DISCUSSIONS

Crowdsourcing contracting entails a best-matching problem, for which a best-matching protocol plays a critical role in evaluating and selecting appropriate design or manufacturing agents for crowdsourcing. Decision tree learning facilitates selection decision making by incorporating past performance data of crowdsourcing agents. This data-driven method suggests good opportunities to extend crowdsourcing contracting to considering a large smart service network over time across a planning horizon.

The emerging business model of crowdsourcing enables the reaching to the external partners through the value creation and capture process. By exploring their knowledge and utilizing their resources, the innovators can focus on their core activities. Meanwhile, the designers and manufacturers can be involved into the product innovation process. The collaborative crowdsourcing faces the challenge from decentralizing and opening of product fulfillment structure. This work focuses on the physical product fulfillment through the crowdsourcing. A product fulfillment flow named collaborative-crowdsourcing product fulfillment has been proposed to entail the crowdsourcing process and the collaboration in this paradigm. This workflow reengineers the conventional product fulfillment process to allow the negotiation and collaboration with external partners.

From the platform-based view, the collaborative-crowdsourcing product fulfillment can be identified as a platform to serve the product fulfillment process. The open innovator, open designers, and open manufacturers can be identified as the modules. The crowdsourcing contracting mechanism and crowdsourcing platform offer an interface for the modules. The formulated workflow of the product fulfillment process provides an embodiment architecture for organizing the modules to an enterprise.

Following this workflow, the contracting mechanism entails the underlying rationale of the tournament-based crowdsourcing. A product fulfillment subtask can be decomposed from an innovation project and allocated by contracting brokers. The contracting brokers hold the tournament to search for the best solutions for the subtasks from the crowds and rewards the corresponding external partners with supply contracts. Their interaction and information exchange are formulated to support the contracting mechanism.

The evaluation mechanism is the backbone of the contracting mechanism since it serves the selection of the best solution. It will not only determine the customer's satisfaction of final products, but also the coordination of the crowds' cohort decision making.

Thanks to the platform theory, various innovators and partners can be rapidly reconfigured to fulfill various customer needs. This research provides the opportunities for the companies to utilize external resources while focusing on their core competitiveness. Specifically, this work eases the SME's innovation and economies of scale.

However, there are several limitations in this work. Firstly, this work only focuses on the contracting stage in the collaborative-negotiation process. After the contracting, the coordination of the design and manufacturing in the execution of collaborative-crowdsourcing product fulfillment needs more research. Secondly, since the collaborative crowdsourcing use open calls to invite the bidding, the intellectual properties protection can be identified as the difficulties. The decentralized product fulfillment requires the protection of the intellectual properties not only during the open calls but also the operation of the reconfigured production. Thirdly, the robustness of the generalized evaluation method needs further exploration.

Several ideas are elaborated for the potential endeavors in the future. Firstly, set-based engineering techniques and predictive best matching protocol should be applied to coordinate the design and manufacturing supply chain in the executive phase. Secondly, the emerging blockchain technologies shed lights on the security of the decentralized networks. Applying blockchain to ensure the intellectual property protection will be an applicable method through collaborative-crowdsourcing product fulfillment. Thirdly, a sensitivity analysis of the evaluation method should be implemented.

REFERENCES

Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. Academy of management review, 37(3), 355-375.

Alexopoulos, C., Seila, A. F., & Banks, J. (1998). Output data analysis. In Handbook of Simulation (No. 7). John Wiley & Sons.

Appleyard, M. M., & Chesbrough, H. W. (2017). The dynamics of open strategy: from adoption to reversion. Long Range Planning, 50(3), 310-321.

Baldwin, C.Y. & Clark, K.B. (2000). Design rules: The power of modularity (Vol. 1).

Banerjee, A. G., Beckmann, B., Carbone, J., DeRose, L., Giani, A., Koudal, P., & Yund, W. (2015, October). Cloud Computing-Based Marketplace for Collaborative Design and Manufacturing. In International Internet of Things Summit (pp. 409-418). Springer, Cham.

Bauwens, M. (2009). The Emergence of Open Design and Open Manufacturing.

Bayus, B. L. (2010). Crowdsourcing and individual creativity over time: the detrimental effects of past success. Available at SSRN 1667101.

Berger, R. (2014) Industry 4.0 - The new industrial revolution - How Europe will succeed, Roland Berger Strategy Consultants.

Bhargava, R., Levalle, R. R., & Nof, S. Y. (2016). A best-matching protocol for order fulfillment in re-configurable supply networks. Computers in Industry, 82, 160-169.

Bogers, M., Zobel, A. K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., ... & Ter Wal, A. L. (2017). The open innovation research landscape: Established perspectives and emerging themes across different levels of analysis. Industry and Innovation, 24(1), 8-40.

Boisseau, É., Omhover, J. F., & Bouchard, C. (2018). Open-design: A state of the art review. Design Science, 4.

Bonabeau, E. (2009). Decisions 2.0: The power of collective intelligence. MIT Sloan management review, 50(2), 45.

Brenner, W., Zarnekow, R., & Wittig, H. (2012). Intelligent software agents: foundations and applications. Springer Science & Business Media.

Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2017). How virtualization, decentralization and network building change the manufacturing landscape: an industry 4.0 perspective. FormaMente, 12.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of ‘big data’. McKinsey Quarterly, 4(1), 24-35.

Brunswicker, S., & Chesbrough, H. (2018). The Adoption of Open Innovation in Large Firms: Practices, Measures, and Risks A survey of large firms examines how firms approach open innovation strategically and manage knowledge flows at the project level. Research-Technology Management, 61(1), 35-45.

Bücheler, T., & Sieg, J. H. (2011). Understanding science 2.0: Crowdsourcing and open innovation in the scientific method. Procedia Computer Science, 7, 327-329.

Capgemini. (2014) Industry 4.0 - The Capgemini Consulting View: Sharpening the Picture beyond the Hype, Capgemini Consulting.

Chesbrough, H., Vanhaverbeke, W., & West, J. (2006). Open innovation. The new Imperative for creating and profiting from technology, 1.

Chesbrough, H. (2006a). Open business models: How to thrive in the new innovation landscape. Harvard Business Press.

Chesbrough, H. (2006). Open innovation: a new paradigm for understanding industrial innovation. Open innovation: Researching a new paradigm, 400, 0-19.

Chesbrough, H. W. (2007). Why companies should have open business models. MIT Sloan management review, 48(2), 22.

Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.). (2006). Open innovation: Researching a new paradigm. Oxford University Press on Demand.

Chesbrough, H., Vanhaverbeke, W., & West, J. (Eds.). (2014). New frontiers in open innovation. Oup Oxford.

Chesbrough, H. W., & Appleyard, M. M. (2007). Open innovation and strategy. California management review, 50(1), 57-76.

Crawford, M. (2018) How Industry 4.0 Impacts Engineering Design, ASME.org.

Dhanaraj, C., & Parkhe, A. (2006). Orchestrating innovation networks. Academy of management review, 31(3), 659-669.

Diener, K., & Piller, F. (2010). The market for open innovation. A Study of the Intermediaries and Brokers for Open Innovation.

Dudek, G., & Stadtler, H. (2005). Negotiation-based collaborative planning between supply chains partners. European Journal of Operational Research, 163(3), 668-687.

ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tseng, M., & Bernard, A. (2013). Product variety management. Cirp Annals, 62(2), 629-652.

ElMaraghy, H., & ElMaraghy, W. (2014). Variety, Complexity and Value Creation. Enabling Manufacturing Competitiveness and Economic Sustainability.

Freitag, M., Becker, T., & Duffie, N. A. (2015). Dynamics of resource sharing in production networks. CIRP Annals, 64(1), 435-438.

Füller, J. (2010). Refining virtual co-creation from a consumer perspective. California management review, 52(2), 98-122.

Gaonkar, R., & Viswanadham, N. (2001). Collaboration and information sharing in global contract manufacturing networks. IEEE/ASME transactions on mechatronics, 6(4), 366-376.

Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. Journal of product innovation management, 31(3), 417-433.

Giannoccaro, I., & Pontrandolfo, P. (2004). Supply chain coordination by revenue sharing contracts. International journal of production economics, 89(2), 131-139.

Govindan, K., & Popiuc, M. N. (2014). Reverse supply chain coordination by revenue sharing contract: A case for the personal computers industry. European Journal of Operational Research, 233(2), 326-336.

Green, D.P., Fuchsberger, V., Kirk, D., Taylor, N., Chatting, D., Meissner, J.L., Murer, M., Tscheligi, M., Lindtner, S., Bjorn, P., Reiter, A. (2017). Open Design at the Intersection of Making and Manufacturing, Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 542-549, Denver, Colorado, ACM New York, NY, ISBN: 978-1-4503-4656-6.

Gupta, A., Agarwal, R. K., & Whitman, L. (2001, December). Supply chain agent decision aid system (SCADAS). In Proceeding of the 2001 Winter Simulation Conference (Cat. No. 01CH37304) (Vol. 1, pp. 553-559). IEEE.

Holland, M., Nigischer, C., & Stjepandić, J. (2017). Copyright protection in additive manufacturing with blockchain approach. In Transdisciplinary Engineering: A Paradigm Shift (pp. 914-921). IOS Press.

Hosseini, M., Phalp, K., Taylor, J., & Ali, R. (2014, May). The four pillars of crowdsourcing: A reference model. In 2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS) (pp. 1-12). IEEE.

Howe, J. (2006). The rise of crowdsourcing. Wired magazine, 14(6), 1-4.

Hu, Q., Che, X., Zhang, L., Zhang, D., Guo, M., & Yu, D. (2011). Rank entropy-based decision trees for monotonic classification. IEEE Transactions on Knowledge and Data Engineering, 24(11), 2052-2064.

Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. Autonomous agents and multi-agent systems, 1(1), 7-38.

Jiang, P., Ding, K., & Leng, J. (2016). Towards a cyber-physical-social-connected and service-oriented manufacturing paradigm: Social Manufacturing. Manufacturing Letters, 7, 15-21.

Jiao, J., Ma, Q., & Tseng, M. M. (2003). Towards high value-added products and services: mass customization and beyond. Technovation, 23(10), 809-821.

Jiao, J. R., Simpson, T. W., & Siddique, Z. (2007). Product family design and platform-based product development: a state-of-the-art review. Journal of intelligent Manufacturing, 18(1), 5-29.

Jiao, J. R., You, X., & Kumar, A. (2006). An agent-based framework for collaborative negotiation in the global manufacturing supply chain network. Robotics and Computer-Integrated Manufacturing, 22(3), 239-255.

Jiao, J., Zhou, F., Gebraeel, N. Z., & Duffy, V. (2020). Towards augmenting cyber-physical-human collaborative cognition for human-automation interaction in complex manufacturing and operational environments. International Journal of Production Research, 58(16), 5089-5111.

Johar, I., Lipparini, F., & Addarii, F. (2015). Making good our future exploring the new boundaries of open & social innovation in manufacturing. Social Innovation Europe,(policy paper).

Kagermann, H., Wahlster, W., & Helbig, J. (Eds.). (2013) Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group.

Kaihara, T. (2001). Supply chain management with market economics. International Journal of Production Economics, 73(1), 5-14.

Kaihara, T., Katsumura, Y., Suginishi, Y., & Kadar, B. (2017). Simulation model study for manufacturing effectiveness evaluation in crowdsourced manufacturing. CIRP Annals, 66(1), 445-448.

Keeney, R. L., Raiffa, H., & Meyer, R. F. (1993). Decisions with multiple objectives: preferences and value trade-offs. Cambridge university press.

Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. Resources, conservation and recycling, 127, 221-232.

Kiulian, A. (2018). Decentralized Artificial Intelligence Is Coming: Here's What You Need To Know, Forbes Technology Council, https://www.forbes.com/sites/forbestechcouncil/2018/01/11/decentralized-artificial-intelligence-is-coming-heres-what-you-need-to-know/#3dbd42f2146d.

Koch, M. D., & Tumer, I. Y. (2009). Towards Open Design: The Emergent Face of Engineering---A Position Paper. In DS 58-3: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 3, Design Organization and Management, Palo Alto, CA, USA, 24.-27.08. 2009 (pp. 97-108).

Kohler, T. (2015). Crowdsourcing-based business models: how to create and capture value. California management review, 57(4), 63-84.

Koomsap, P. (2013). Design by customer: concept and applications. Journal of Intelligent Manufacturing, 24(2), 295-311.

Kortmann, S., & Piller, F. (2016). Open business models and closed-loop value chains: Redefining the firm-consumer relationship. California Management Review, 58(3), 88-108.

Koufteros, X., Lu, G., Peters, R. C., Lai, K. H., Wong, C. W., & Cheng, T. E. (2014). Product development practices, manufacturing practices, and performance: A mediational perspective. International Journal of Production Economics, 156, 83-97.

Kusiak, A. (2019). Service manufacturing: Basic concepts and technologies. Journal of Manufacturing Systems, 52, 198-204.

Lazzarotti, V., & Manzini, R. (2009). Different modes of open innovation: a theoretical framework and an empirical study. International journal of innovation management, 13(04), 615-636.

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing letters, 3, 18-23.

Leimeister, J. M., Huber, M., Bretschneider, U., & Krcmar, H. (2009). Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition. Journal of management information systems, 26(1), 197-224.

Li, Z., Wang, W. M., Liu, G., Liu, L., He, J., & Huang, G. Q. (2018). Toward open manufacturing: A cross-enterprises knowledge and services exchange framework based on blockchain and edge computing. Industrial Management & Data Systems.

Lüttgens, D., Pollok, P., Antons, D., & Piller, F. (2014). Wisdom of the crowd and capabilities of a few: internal success factors of crowdsourcing for innovation. Journal of Business Economics, 84(3), 339-374.

Malhotra, A., & Majchrzak, A. (2014) Managing Crowds in Innovation Challenges.

Mansouri, S. A., Gallear, D., & Askariazad, M. H. (2012). Decision support for build-to-order supply chain management through multiobjective optimization. International Journal of Production Economics, 135(1), 24-36.

Mattsson, S., Fasth, Å., Berlin, C., & Stahre, J. (2012). Describing Human-Automation Interaction in production. In Swedish Production Symposium, SPS12.

Maurer, S., & Scotchmer, S. (2006). Open source software: the new intellectual property paradigm.

Meyer, M. H., & Lehnerd, A. P. (1997). The Power of Product Platforms: Building Value and Cost Leadership, 1997. New York, NY, 10020, 39.

Mingers, J. (1989). An empirical comparison of selection measures for decision-tree induction. Machine learning, 3(4), 319-342.

Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., ... & Ueda, K. (2016). Cyber-physical systems in manufacturing. Cirp Annals, 65(2), 621-641.

Pan, Z., Yu, H., Miao, C., & Leung, C. (2016, March). Efficient collaborative crowdsourcing. In Thirtieth AAAI Conference on Artificial Intelligence.

Pine, B. (2009). The Future of Mass Customization. Special Talk in Mass Customization and Personalization Conference (MCPC2009).

Pine, B.J. (1993). Mass customization: the new frontier in business competition: Harvard Business Press.

Potharst, R., Ben-David, A., & van Wezel, M. (2009). Two algorithms for generating structured and unstructured monotone ordinal data sets. Engineering Applications of Artificial Intelligence, 22(4-5), 491-496.

Quinlan, R. (1993), C4.5: Programs for Machine Learning. Morgan Kaufmann.

Raasch, C., Herstatt, C., & Balka, K. (2009). On the open design of tangible goods. R&d Management, 39(4), 382-393.

Ragin, C. C., Drass, K. A., & Davey, S. (2006). Fuzzy-set/qualitative comparative analysis 2.0. Tucson, Arizona: Department of Sociology, University of Arizona, 23(6), 1949-1955.

Ratto, M., & Ree, R. (2012). Materializing information: 3D printing and social change. First Monday.

Redlich, T., & Bruhns, F. L. (2008). Open production: a new broker-based approach to interactive value creation and user manufacturing. In ASME International Mechanical Engineering Congress and Exposition (Vol. 48654, pp. 181-189).

Richardson, M. (2016). Pre-hacked: Open Design and the democratisation of product development. New Media & Society, 18(4), 653-666.

Rosen, P. A. (2011). Crowdsourcing lessons for organizations. Journal of Decision Systems, 20(3), 309-324.

Rosen, P. A. (2011). Decision Making in a Web 2.0 Environment. Crowdsourcing Lessons for Organizations. Journal of Decision Systems, 20, 309-324.

Rosen, R., Von Wichert, G., Lo, G., & Bettenhausen, K. D. (2015). About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine, 48(3), 567-572.

Sadeh, N. M., Hildum, D. W., Kjenstad, D., & Tseng, A. (2001). MASCOT: an agent-based architecture for dynamic supply chain creation and coordination in the internet economy. Production Planning & Control, 12(3), 212-223.

Sahin, F., & Robinson Jr, E. P. (2005). Information sharing and coordination in make-to-order supply chains. Journal of operations management, 23(6), 579-598.

Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. Information Systems Management, 30(1), 2-20.

Schaefer, D. (2014), Cloud-Based Design and Manufacturing (CBDM): A Service-Oriented Product Development Paradigm for the 21st Century, ISBN 978-3-319-07398-9, Springer.

Schuh, G., Potente, T., Wesch-Potente, C., Weber, A. R., & Prote, J. P. (2014). Collaboration Mechanisms to increase Productivity in the Context of Industrie 4.0. Procedia Cirp, 19, 51-56.

Shin, M., & Jung, M. (2004). MANPro: mobile agent-based negotiation process for distributed intelligent manufacturing. International journal of production research, 42(2), 303-320.

Simard, C., & West, J. (2006). Knowledge networks and the geographic locus of innovation. Open innovation: researching a new paradigm, 220-240.

Subramanian, R., Gupta, S., & Talbot, B. (2009). Product design and supply chain coordination under extended producer responsibility. Production and Operations Management, 18(3), 259-277.

Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. Economies, Societies and Nations, 296.

Swaminathan, J. M., Smith, S. F., & Sadeh, N. M. (1998). Modeling supply chain dynamics: A multiagent approach. Decision sciences, 29(3), 607-632.

Thurston, D. L., & Carnahan, J. V. (1992). Fuzzy ratings and utility analysis in preliminary design evaluation of multiple attributes.

Thurston, D. L. (1991). A formal method for subjective design evaluation with multiple attributes. Research in engineering Design, 3(2), 105-122.

Tseng, M.M., Jiao, J., & Merchant, M.E. (1996). Design for mass customization. CIRP Annals-Manufacturing Technology. 45:153-6.

Tseng, M. M., Jiao, R. J., & Wang, C. (2010). Design for mass personalization. CIRP annals, 59(1), 175-178.

Vallance, R., Kiani, S., & Nayfeh, S. (2001, May). Open design of manufacturing equipment. In Proceedings of the CHIRP 1st International Conference on Agile, Reconfigurable Manufacturing (pp. 33-43).

Van de Vrande, V., De Jong, J. P., Vanhaverbeke, W., & De Rochemont, M. (2009). Open innovation in SMEs: Trends, motives and management challenges. Technovation, 29(6-7), 423-437.

Von Neumann, J., & Morgenstern, O. (1945). Theory of games and economic behavior. Bull Amer Math Soc. 51:498-504.

Weber, S. (2004). The success of open source: Harvard University Press.

Weyer, S., Schmitt, M., Ohmer, M., & Gorecky, D. (2015). Towards Industry 4.0-Standardization as the crucial challenge for highly modular, multi-vendor production systems. Ifac-Papersonline, 48(3), 579-584.

Wladawsky-Berger, I. (2018). Blockchain and the Promise of an Open, Decentralized Internet. Wall Street Journal CIO blog, February, 23.

Wu, D., Rosen, D. W., Wang, L., & Schaefer, D. (2015). Cloud-based design and manufacturing: A new paradigm in digital manufacturing and design innovation. Computer-Aided Design, 59, 1-14.

Wu, D., Schaefer, D., & Rosen, D. W. (2013a). Cloud-based design and manufacturing systems: a social network analysis. In DS 75-7: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, Vol. 7: Human Behaviour in Design, Seoul, Korea, 19-22.08. 2013.

Wu, D., Thames, J. L., Rosen, D. W., & Schaefer, D. (2013). Enhancing the product realization process with cloud-based design and manufacturing systems. Journal of Computing and Information Science in Engineering, 13(4).

Xiao, T., Qi, X., & Yu, G. (2007). Coordination of supply chain after demand disruptions when retailers compete. International Journal of Production Economics, 109(1-2), 162-179.

Xu, P. J., Allgor, R., & Graves, S. C. (2009). Benefits of reevaluating real-time order fulfillment decisions. Manufacturing & Service Operations Management, 11(2), 340-355.

Xu, X. (2012). From cloud computing to cloud manufacturing. Robotics and computer-integrated manufacturing, 28(1), 75-86.

Zhang, L., Luo, Y., Tao, F., Li, B. H., Ren, L., Zhang, X., ... & Liu, Y. (2014). Cloud manufacturing: a new manufacturing paradigm. Enterprise Information Systems, 8(2), 167-187.

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. Information Systems Frontiers, 16(3), 417-434.

Zhen, L., Jiang, Z., & Song, H. T. (2011). Distributed knowledge sharing for collaborative product development. International Journal of Production Research, 49(10), 2959-2976.

DECISION SCIENCES INSTITUTE

Assessing Injury Severity in Passenger Vehicle Crashes: A Case of Data Balancing, Automatic Feature Selection & Machine Learning

Rupesh Kumar Agrawal
Northern Kentucky University, USA
Email: agrawalr1@nku.edu

Robert Fritts
Oklahoma State University, USA
Email: robert.fritts@okstate.edu

Dursun Delen,
Oklahoma State University, USA
Email: dursun.delen@okstate.edu

ABSTRACT

The Year 2016 was the deadliest year on American roads, with more than 40,000 deaths, an increase of 6% over the Year 2015. In 2019, the average cost of all motor-vehicle crashes was estimated at $11,880,000. The current study compares manual and automatic feature selection to identify the best set of factors to predict and explain post-accident injury severity. Contrary to prevalent belief, the under-sampling method performed better as compared to ordinary over-sampling (SMOTE). In addition, developed models such as the Decision Tree, Random Forest, and Neural Network to derive high prediction accuracy and intuitive insight.

KEYWORDS: Automotive, Crash, Injury Severity, Data Balance, Feature Selection, Prediction, and Machine Learning

INTRODUCTION

Automotive crashes cause a significant burden on our economy and society at large. Crashes lead to lost productivity, workplace losses, legal expenses, medical costs, emergency medical services, insurance administration costs, congestion costs, and property damage costs (NHTSA, 2009). In 2019, the average cost estimated based on per death for all motor-vehicle crashes (fatal, nonfatal injury, and property damage) was $11,880,000 (National Safety Council, 2021). In addition to financial costs, the human toll is exceptionally high. For example, 2016 was the deadliest year on American roads, with more than 40,000 deaths, an 6% increase over 2015 (NHTSA, 2017). In addition, motor vehicle accidents were the leading cause of death for children aged ten and young people 16 to 23 in 2015 (NHTSA, 2015).

The National Highway Traffic Safety Administration (NHTSA) leads research on vehicle and behavioral safety as part of the federal Department of Transportation (DOT). NHTSA's mission is to strategize, plan, and implement research to reduce crashes, fatalities, and injuries to support Congressional Mandates, DOT & NHTSA goals. National Automotive Sampling System (NASS), Fatality Analysis Reporting System (FARS), and General Estimates System (GES) are programs that develop census and statistically valid sample data of crashes within the US. National Automotive Sampling System (NASS) is composed of two systems - the Crashworthiness Data System (CDS) and the General Estimates System (GES). The NASS GES compiles probability-based samples of police-reported crashes representative of sixty locations and more than five million reported accidents – this NASS GES data is the starting point for this study.

With support from insurance companies and the government, the automotive industry has come a long way in understanding causes and measures to reduce accidents. As a result, the mortality rate has decreased per mile traveled. However, miles traveled are grown. As a result, the total death count has risen (NHTSA, 2016). Prevention is always better than cure. So, it is vital to continually enhance understanding to reduce injury and fatality rates as traveled miles increase.

In a 2005 study, researchers concluded that understanding probable factors (variables) that may help reduce injury is key to preventing accidents (Delen et al., 2006). NHTSA has recorded extensive data on accident and injury severity factors such as demographic data (age, gender), behavioral characteristics (use of drugs or alcohol), accident conditions (number of occupants, distractions present), and environmental factors such as weather and roadway conditions. Definition and factors of accident and injury severity are evolving, consequently understanding factors are important to understand causes of total mortality.

The current study followed the cross-industry process for data mining (CRISP-DM) methodology (Wirth & Hipp, 2000). CRISP-DM method guided steps from developing study goals for evaluation of the results. Also, guided establishing two objectives for passengers injured in vehicle accident: 1) identify best predictors of an incapacitating or fatal injury versus a minor injury, and 2) predict post-accident injury severity based on identified factors from the first objective.

The current study expands the work on injury severity (Delen et al., 2006) by including elements previously not covered. The study developed research questions that formed the basis of added data mining objectives. The current study shall contribute to the existing work by informing on the best methodologies for analyses to support the following questions: 1) Can machine learning (ML) techniques be applied for automatic variable selection? 2) What sampling (data balancing) techniques are most useful? 3) Are any sampling performance gains worth the computational cost?

LITERATURE REVIEW

Research studies have developed extant literature primarily to understand the dynamics of accidents, factors contributing to accidents, and accident-related injury severities. Few studies have focused on factors such as distraction (texting, talking), and some studies are specific on factors such as use of alcohol as the cause of the accident. In other approaches, research work has focused on decoding the complexity of passenger vehicles. For example, the influence of vehicle factors (occupant contacts, roof damage) on occupant injury frequency, and the effects of occupant ejection from rolled-over cars crashes and had top damage from ground contact (Huelke et al., 1983). Yu et al. (2014) explored injury severity on high-speed roads with real-time traffic data. In 2015, Haleem et al. explored pedestrian injury severity at the signal and non-signal intersections. Finally, Uddin et al. (2017) studied trucks crash injury severity due to road lighting conditions. The above-cited literature/body of knowledge highlights the rich research work and explores vehicle crash injury severity using data mining/analytics methods for finding unique factors.

In addition, contemporary research has focused on identifying best machine learning algorithms (linear & non-linear) to test and support hypotheses relating to accident-related injury severity. For example, Mussone et al. (1999) used artificial neural networks (ANN) to develop models for vehicular accidents in Milan. The authors justified the choice of ANN over linear models' citing inferior statistical ability of linear models to generate robust models. Similarly, Yu et al. (2014), used Random Forest (RF) models to rank variable importance. RF models have been often used to rank variable importance in traffic safety studies to provide unbiased error estimates and avoid cross-validation (Breiman et al., 2000). In another study, Chen et al. (2016) used Decision Tree (DT) models to explain injury severity of drivers because DT is easy to understand and is easy to explain.

The above literature review favors the argument for the current study goals, models (DT, ANN, and Ensemble RF), and RF for rank selection for variable (factor)selection. The compiled study results help compare model performance using the confusion matrix, the area under the curve (AUC), and sensitivity/specificity measures.

MODEL

CRISP-DM methodology guided the steps to refine, finalize the business and data mining goals, proceeded with data processing, modeling, model evaluation, and objective assessment. CRISP-DM methodology influences the overarching goals of identifying variables that contribute to injury severity, developing a rich dataset to support a confident model, and selecting the best model from the applied models. Figure 1 shows the overall study process.

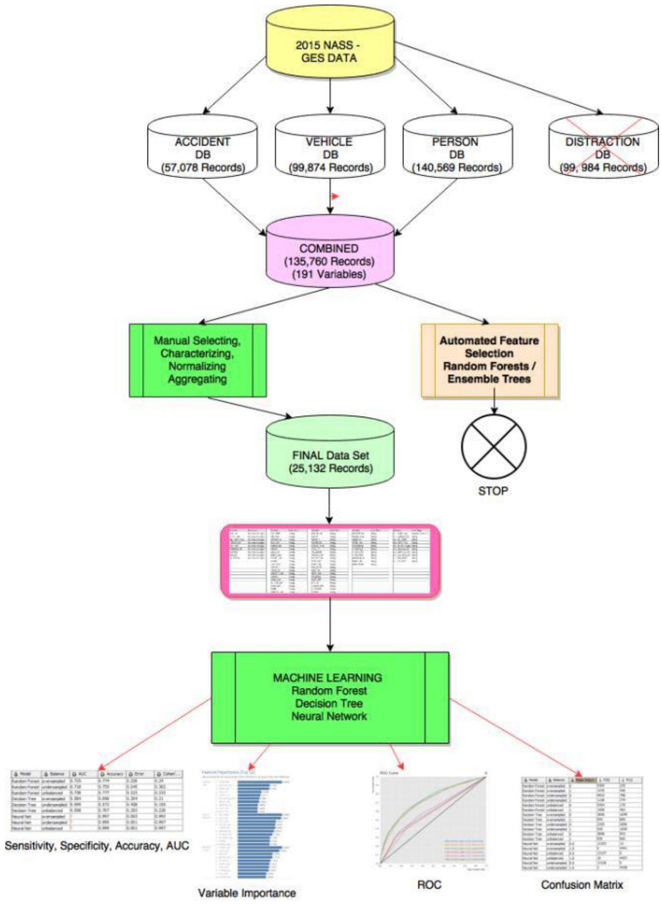


Figure 1: Overall Process-Based on CRISP-DM Methodology

The next section of the study is conducted in the following steps: Data Understanding, Data

Preparation, Modeling, and Evaluation. The study required 70-80% time in data understanding and preparation and 10-20% time in model building, evaluation, and compilation of results.

Dataset from NASS GES (2015) initiated the study. Initial data analysis of the NASS GES dataset and motivation from extant literature guided the selection of the target population for the study. It was clear from the fatalities data from FARS (Refer to Figure 2) that passenger cars are the most affected groups from accidents. Consequently, the target population for the study are the passengers with car accidents. The area of inquiry is further refined to segregate best predictors of the accidents by classifying the severity of the post-accident injury (incapacitating or fatal injury versus a minor injury). Above accident scenario (passenger car), and variables (fatal vs. minor) helped develop the final predication models.

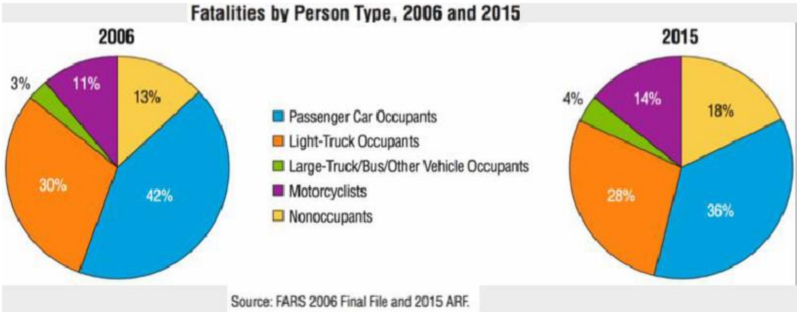


Figure 2: Fatalities by Person Type, 2006 & 2015

The growth in technology, reduction in cost, global availability, and aggressive development in machine learning algorithms, and the need to conduct a just-in-time task have increased dependence on computers. As a result, computers are enabling automated analytics work from variable selection to final deployment. Conventional data mining projects start with data exploration to select a few target variables to optimize the overall data mining effort. However, with the popularity/availability of big data and fast processing power, techniques have been developed to automate the data mining processes. For example, Principal Component Analysis (PCA) and Random Forest methods limit human interaction to an ideal level and support the best model development for production.

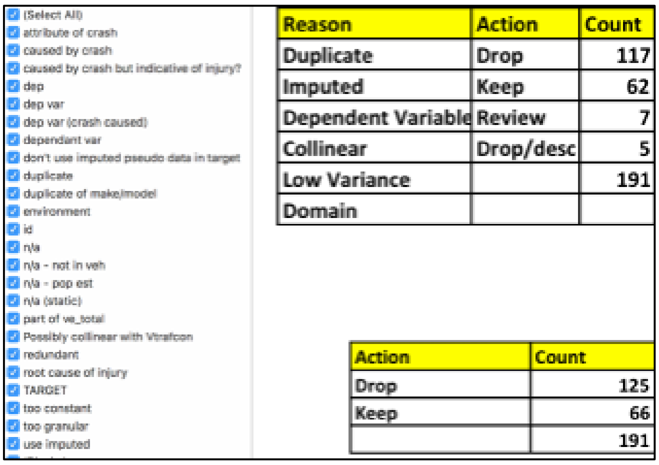
In addition, extant and contemporary works are available for insight and techniques for data dimensionality reduction (KNIME, 2015; Genuer, 2010; Burger, 2011). The above literature established following data mining goals; a) Identify variables and their relative importance, b) Identify the best model for implementation, and c) Establish validity of automated variable selection via machine learning techniques.

Data Understanding & Preparation

The study acquired the raw data from the NASS GES website (GES, 2015). Raw data included four related tables from the database: VEHICLE, PERSON, ACCIDENT, and DISTRACT. The VEHICLE dataset lists the in-transport motor vehicle, driver, and pre-crash data. Refer to Figure 3 below, showing the left join between the ACCIDENT and VEHICLE data file and the inner join between VEHICLE data and PERSON data file. To support objective the DISTRACT table was excluded from the study. Once data were combined, the research team conducted several iterations to select variables for the final model run.

Figure 3: Venn Diagram | Data Merge

The manual variable selection method utilized extensive literature studies and commonsense for variable selection. For example, variables such as alcohol content in the driver may contribute to better prediction results. In addition, crash data review helped identify duplicate columns, imputed variables, dependent variables, variables collinear with other variables, and lastly, variables with low variance (KNIME, 2015). Refer to Figure 4, showing the process of selecting the variables using the pivot table and the final count of eliminated and selected variables.



Reason	Action	Count
Duplicate	Drop	117
Imputed	Keep	62
Dependent Variable	Review	7
Collinear	Drop/desc	5
Low Variance		191
Domain		

Action	Count
Drop	125
Keep	66
	191

Figure 4: Data Understanding | Collect | Explore | Describe

GES crash data is grouped into three main stratums: passenger vehicle, medium & heavy-duty truck, and others. Refer to Figure 5, showing the distribution of each vehicle type.

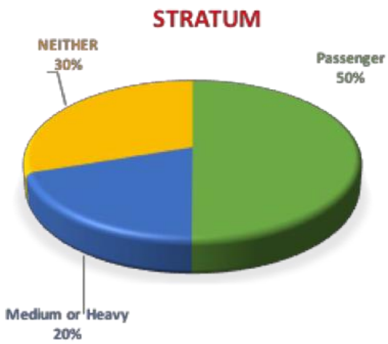


Figure 5: Data Record Distribution Based on Vehicle Type

In this part of the study, data cleaning using transformation and reduction using various techniques available in KNIME helped consolidate the data. First, data were transformed to manage nulls, aggregate, and reduce range. Use of RULE ENGINE NODES helped create nine dimensions. MATH FORMULA NODE aided in deriving the vehicle model age (AGE = YEAR – MODYR_IM). The current step used relevant nodes to filter, visualize, and explore the newly derived variables without introducing (data) defects, artifacts, nulls, or otherwise. In the second step, we transformed all the variables into relevant data types, i.e., numeric, categorical, ordinal or nominal, and discrete or continuous. Due to the extensive transformations, particularly from numeric to string, each variable's domain (range) was reestablished using a DOMAIN CALCULATORnode with correct range information (list of values).

Table 1: Injury Severity Breakdown

DESCRIPTION	INJ_SEVERITY	COUNT	TOTAL
No Injury (O)	0	41,101	41,101
Possible Injury (C)	1	11,317	25,132
Non-incapacitating Evident Injury (B)	2	7,449	
Incapacitating Injury (A)	3	5,665	
Fatal Injury (K)	4	701	
Injured, Severity Unknown (U)	5	709	2,057
Died Prior to Crash	6	1	
Unknown	9	1,347	

Modeling

Modeling is conducted in two processes. The first process predicts variable importance automatic feature selection. Random Forest (RF) learner was used with high iteration (3,000) and a very shallow depth (3), and the result was then passed to the predictor node to ensure model would successfully test. This section of the study was setup to confirm the use of automatic machine learning (ML) techniques for the successful prediction of variable importance compared to the manual variable selection and model development. The primary differentiator of this section (automatic feature) is that it uses a single ML model i.e., RF. And, while the data set is cleaned, data is not run through dimensionality reduction applied to the manual modeling section. Use of the current step allows us to compare dimensionality reduction techniques for suitability of an automatic method for factor (variable) selection.

The data processing included dimension (variable) reduction using automatic filters available in KNIME to screen out invariant variables, unique IDs, and collinear variables prior to model pre-processing, training, and assessment. The total of 201 variables resulted in 105 variables. These 105 variables were then normalized and ranked based on split selection percentage. the top twenty (20) shown in Figure 6.

In addition to removing factors based on correlation and variance, the second process involved manual exclusion (or selection) of factors based on domain research or knowledge. Consequently, the manual dimension selection reduced 201 variables to sixty-nine (69) final factors. The final sixty-nine (69) factors are noted in Table 2. Note that first process resulted in 105 factors compared to 69 factors in the second process.

Figure 6: Variable Selection with Importance

This next step is common and applied to both automatic and manual variable selection. After processing and transforming data, stratified sampling is applied to partition the data into 70% for training and 30% for testing. In addition to unbalanced data, *under-sampling* and Synthetic Minority Oversampling Technique *(SMOTE)* are applied to compare the impact of balancing methods. Extant research has preferred SMOTE (over-sampling) method as superior to the under-sampling method (Chawla et al., 2002; Fernández et al., 2018).

Table 2: Final List of Variables

Feature	Data Type	Feature	Data Type	Feature	Data Type	Feature	Data Type	Feature	Data Type
AGE_IM	Number (integer)	ACC_TYPE	String	MDLYR_IM	String	VEVENT_IM	String	der_MOD_Age	Number (integer)
ALCHL_IM	Number (integer)	AIR_BAG	String	MODEL	String	VNUM_LAN	String	der_AIRBAG_YN	String
der_MOD_Age	Number (integer)	BDYTYP_IM	String	MONTH	String	VPROFILE	String	der_DY_PART	String
HITRUN_IM	Number (integer)	BUS_USE	String	NUMOCCS	String	VSPD_LIM	String	der_EJECT_YN	String
INT_HWY	Number (integer)	CARGO_BT	String	PCRASH1_IM	String	VSURCOND	String	der_INJSEV_major	String
PERALCH_IM	Number (integer)	DRUGS	String	PER_TYP	String	VTCONT_F	String	der_RestUse_YN	String
PERMVIT	Number (integer)	EJECT_IM	String	REL_ROAD	String	VTRAFCON	String	der_RollOver_YN	string
SEX_IM	Number (integer)	EMER_USE	String	RELJCT1_IM	String	VTRAFWAY	String	der_SpeedRel_YN	String
VE_TOTAL	Number (integer)	EVENT1_IM	String	RELJCT2_IM	String	WEATHR_IM	String	der_WK_PART	String
		GVWR	String	REST_MIS	String	WKDY_IM	String	der_YR_PART	String
		HAZ_CNO	String	REST_USE	String	WRK_ZONE	String		
		HAZ_REL	String	ROLLOVER	String				
		HOUR_IM	String	SEAT_IM	String				
		IMPACT1_IM	String	SPEC_USE	String				
		J_KNIFE	String	SPEEDREL	String				
		LAND_USE	String	TOW_VEH	String				
		LGTCON_IM	String	TYP_INT	String				
		LOCATION	String	V_ALCH_IM	String				
		MAKE	String	V_CONFIG	String				
		MANCOL_IM	String	VALIGN	String				

Results

The following outputs were consolidated to measure and compare the performance of each model; 1- Variable Importance (Figure 7), 2-Joined predictions for blended ROC curves (Figure 7), 3-Consolidated confusion matrix stats (Table 3), and 4-Consolidated accuracy & ROC AUC stats (Figure 8).

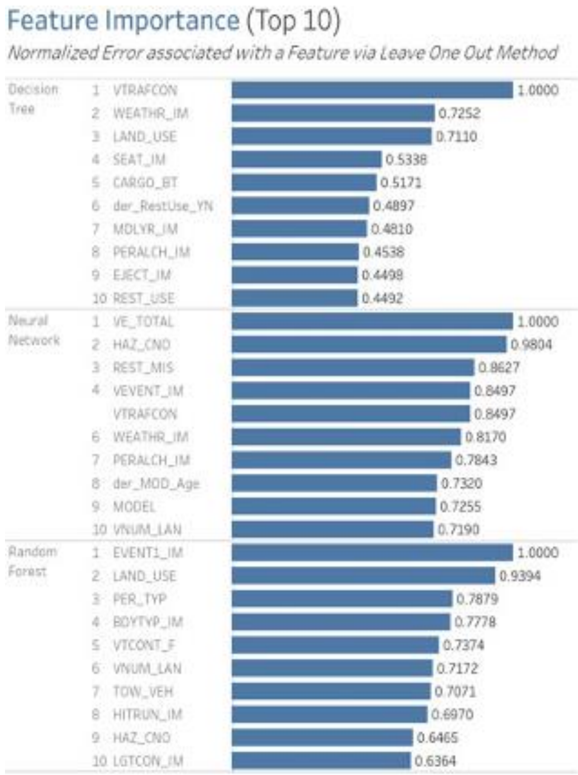
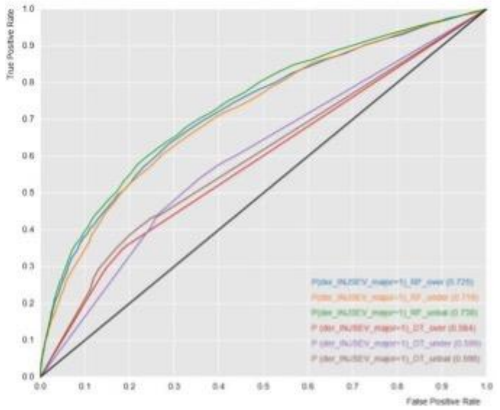


Figure 7: Variable Selection with Importance

Table 3: Confusion Matrix

Model	Balance	Major Injury	P(0)	P(1)
Random Forest	oversampled	0	5395	235
Random Forest	oversampled	1	1470	440
Random Forest	undersampled	0	4922	708
Random Forest	undersampled	1	1136	774
Random Forest	unbalanced	0	5454	176
Random Forest	unbalanced	1	1508	402
Decision Tree	oversampled	0	3656	1049
Decision Tree	oversampled	1	844	669
Decision Tree	undersampled	0	2355	2006
Decision Tree	undersampled	1	525	1030
Decision Tree	unbalanced	0	3698	953
Decision Tree	unbalanced	1	856	666
Neural Net	oversampled	0.0	13103	13
Neural Net	oversampled	1.0	5	4441
Neural Net	unbalanced	0.0	13137	0
Neural Net	unbalanced	1.0	10	4437
Neural Net	undersampled	0.0	13136	0
Neural Net	undersampled	1.0	3	4438

Due to the difference in Neural Network outputs, the ROC curve is not applicable and has been removed. In addition, the confusion matrix output also had a noise that had to be filtered out, which was unique to the NN model (Refer to Figure 8, Column titled "D AUC").



S Model	S Balance	D AUC	D Accuracy	D Error	D Cohen'...
Random Forest	oversampled	0.725	0.774	0.226	0.24
Random Forest	undersampled	0.718	0.755	0.245	0.302
Random Forest	unbalanced	0.738	0.777	0.223	0.233
Decision Tree	oversampled	0.584	0.696	0.304	0.21
Decision Tree	undersampled	0.599	0.572	0.428	0.155
Decision Tree	unbalanced	0.598	0.707	0.293	0.228
Neural Net	oversampled	?	0.997	0.003	0.993
Neural Net	undersampled	?	0.999	0.001	0.997
Neural Net	unbalanced	?	0.999	0.001	0.997

Figure 8: ROC, Accuracy, Error, and Cohen's Kappa

All models ran successfully. Decision Tree (DT) underperformed with a maximum accuracy of 70%, followed by Random Forest (RF) with improved accuracy of 77%. However, Neural Network performance was too good to be confirmed with accuracy at 99% or better.

In all cases, the unbalanced training set and resulting model outperformed smote and under-sampled method. The result is significant as the balancing techniques come at a high computational cost. On the other hand, variable selection with automatic ML techniques was less successful, showing a little or no consistency between automatic and manual ML modeling methods.

DISCUSSION AND CONCLUSIONS

SMOTE or over-sampling method had the lowest performance. The result is confirmed by researcher Blagus & Lusa (2013) that for high-dimensional data, SMOTE produces desired results in k-NN based models and biases the classification towards the minority class. Additionally, it is expected that balancing heavily categorical data sets would not be productive since balancing relies on mathematical techniques that do not apply well to non-numerical data.

Overall results were not satisfactory per our first hypothesis. Using sixty-nine (69) variables, the authors expected a better accuracy and considerable overlap of variable importance between manual and automatic feature selection. With accuracies DT (70%) & RF (77%) and suspiciously high accuracy of NN (99%), further investigation is required. Nevertheless, the study did favor results towards data mining goals. The Random Forest Ensemble model for variable importance ran successfully. Although the results were not used to develop models, authors believe further investigation towards manual variable selection is called to improve the results. In addition, the model showed better results with equal sampling vs. SMOTE, showing an opportunity to improve computing performance in high-dimensional data modeling. Lastly, we believe exploring ensemble models shall bring added insight for improving model performance.

References (Bibliography)

Amarasingha, Niranga, and Sunanda Dissanayake. "Modeling injury severity of young drivers using highway crash data from Kansas." Journal of the transportation research forum. Vol. 52. No. 1. Transportation Research Forum, 2013.

Anderson, Jason C., and Shangjia Dong. "Heavy-Vehicle Driver Injury Severity Analysis by Time of Week: A Mixed Logit Approach Using HSIS Crash Data." Institute of Transportation Engineers. ITE Journal 87.9 (2017): 41.

Anderson, Jason C., and Shangjia Dong. "Heavy-Vehicle Driver Injury Severity Analysis by Time of Week: A Mixed Logit Approach Using HSIS Crash Data." Institute of Transportation Engineers. ITE Journal 87.9 (2017): 41.

Breiman, Leo. Some infinity theory for predictor ensembles. Technical Report 579, Statistics Dept. UCB, 2000.

Burger, James, and Aoife Gowen. "Data handling in hyperspectral image analysis." Chemometrics and Intelligent Laboratory Systems 108.1 (2011): 13-22.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

Chen, Cong, et al. "An explanatory analysis of driver injury severity in rear-end crashes using a decision table/Naïve Bayes (DTNB) hybrid classifier." Accident Analysis & Prevention 90 (2016): 95-107.

Delen, Dursun, et al. "Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods." Journal of Transport & Health 4 (2017): 118-131.

Delen, Dursun, Ramesh Sharda, and Max Bessonov. "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks." Accident Analysis & Prevention 38.3 (2006): 434-444.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. Journal of artificial intelligence research, 61, 863-905.

Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. "Variable selection using random forests." Pattern Recognition Letters 31.14 (2010): 2225-2236.

GES, 2015. National Automotive Sampling System General Estimates System. NASS GES.

Guide to Calculating Costs - Data Details - Injury Facts. (2021, January 27). Retrieved May 31, 2021, from Injury Facts website: https://injuryfacts.nsc.org/all-injuries/costs/guide-to-calculating-costs/data-details/#:~:text=Expressed%20on%20a%20per%20death,crashes%20(including%20minor%20injuries).

Haleem, Kirolos, Priyanka Alluri, and Albert Gan. "Analyzing pedestrian crash injury severity at signalized and non-signalized locations." Accident Analysis & Prevention 81 (2015): 14-23.

Hayes, John R., and Jonathan I. Groner. "Using multiple imputation and propensity scores to test the effect of car seats and seat belt usage on injury severity from trauma registry data." Journal of pediatric surgery 43.5 (2008): 924-927.

Huelke, Donald F., and Charles P. Compton. "Injury frequency and severity in rollover car crashes as related to occupant ejection, contacts and roof damage: —An analysis of national crash severity study data—." Accident Analysis & Prevention 15.5 (1983): 395-401.

Blagus, R., Lusa, L. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 14, 106 (2013). https://doi.org/10.1186/1471-2105-14-106

Mussone, Lorenzo, Andrea Ferrari, and Marcello Oneta. "An analysis of urban collisions using an artificial intelligence model." Accident Analysis & Prevention 31.6 (1999): 705-718.

National Highway Traffic Safety Administration. "National Automotive Sampling System (NASS) General Estimates System (GES) Analytical User's Manual 1988-1999." US Department of Transportation (2000).

National Highway Traffic Safety Administration. "Traffic safety facts: 2007 data: pedestrians." Annals of Emergency Medicine53.6 (2009): 824.

National Safety Council, 2015. NSC Motor Vehicle Fatality Estimates. National Safety Council, Itasca, Illinois, Last accessed on November 5, 2015 at http:// www.nsc.org/NewsDocuments/2015/6-month- fatality-increase.pdf.

Tech, Knime Blog. "Seven Techniques for Data Dimensionality Reduction." Https://Www.knime.com/Blog/Seven-Techniques-for-Data-Dimensionality-Reduction, KNIME AG, 5 Dec. 2015, 12:38, www.knime.com.

Uddin, Majbah, and Nathan Huynh. "Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways." Accident Analysis & Prevention 108 (2017): 44-55.

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining (Vol. 1). London, UK: Springer-Verlag.

Yu, Rongjie, and Mohamed Abdel-Aty. "Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data." Accident Analysis & Prevention 62 (2014): 161-167

DECISION SCIENCES INSTITUTE
Performance implications of ISO/IEC 27001 certification

Matteo Podrecca
University of Udine
Email: podrecca.matteo@spes.uniud.it

Giovanna Culot
University of Udine
Email: culot.giovanna@spes.uniud.it

Guido Nassimbeni
University of Udine
Email: guido.nassimbeni@uniud.it

Marco Sartor
University of Udine
Email: marco.sartor@uniud.it

ABSTRACT

The purpose of this paper is to explore the operational impact of the adoption of the most renowned norm in the field of information security: ISO/IEC 27001. We develop six research hypotheses; three of them related to firm's operating performance and three which shed light on the moderating role of some contextual factors. The results indicate that the ISO/IEC 27001 certification improves the profitability and the labor productivity of the adopting firms while no effect is recorded on sales performance. The impact appears affected by the munificence of the industry and the level of internationalization of the firm.

KEYWORDS: ISO/IEC 27001, ISO 27001, information security, management system standards, event study

DECISION SCIENCES INSTITUTE

Crisis-inspired Innovation: The role of Innovation-based Identification and Covid-19 Reponse

My-Trinh Bui
VNU International School - Vietnam National University, Hanoi, Vietnam
Email: trinhbm@isvnu.vn

Thi-Thanh-Huyen Tran
Banking Academy, Vietnam
Email: huyenttt@hvnh.edu.vn

Hong-Minh Phi
Vietnam Academy of Social Sciences
Email: minhphihong@gmail.com

Ngoc-Dung Nguyen
Hanoi Metropolitan University
Email: nndung@daihocthudo.edu.vn

ABSTRACT

The study proposes a crisis-inspired innovation framework under the aspect of the knowledge-based view. We initiate knowledge transfer mode based on knowledge characteristics as ambiguity (innovation-based identification mode) and context-dependence (pandemic-response mode). First, the study initiates by figuring out some crucial aspects of innovation-based identification including its influence on innovative culture, corporative culture, and adaptation. Innovation-based identification is expected to be the dominant dimension generating significant crisis-inspired innovation. Second, the pandemic response is expected to accelerate higher levels of organizational innovation amid the COVID-19 outbreak. The determinants of firm response are including magnitude, controllability, status certainty, technology speed.

KEYWORDS: Crisis-inspired Innovation, Identification, Covid-19 Response

INTRODUCTION

The Chinese word "crisis" is made up of two characters, one representing threat and the other representing opportunity. The COVID-19 pandemic virtually put every part of the business into disarray, especially the way businesses engage with their clients, customers, and the way supply chains deliver goods and services. COVID-19 radically transform the way people conduct business, with just as many claiming the situation would have a long-term effect on their customers' needs (Am et al., 2020). Because of unprecedented crisises and events in the environments, the firm must adapt to a rapidly changing market. It is difficult to predict and quantify changes that occur at predicted rates of velocity and size (Brosseau et al., 2019). Many studies have found that dynamic capacities (Colombo et al., 2020), economic factors (Fainshmidt et al., 2016), institutional conditions (Dunning and Lundan, 2009), environmental dynamism (Karna et al., 2016) are the most significant influencing factors on firm performance. Furthermore, the environmental crisis can provide an opportunity to acquire new dynamic capabilities through creative solutions (Ebersberger and Kuckertz, 2021). Aside from COVID-19 health concerns, innovation has evolved into a means of responding to societal and business shifts. Many new businesses arise to capitalize on the expanding entrepreneurial opportunities (Kuckertz et al., 2020). Other existing businesses restructure their operations and launch new approaches and innovations (Kraus et al., 2020).

Although the organizational awareness literature indicates a number of possible explanations for firms' failure to react to environmental externalities, such as prejudice against the task environmental change (e.g., (Elenkov, 1997), bounded rationality (e.g., (Lejarraga and Pindard-Lejarraga, 2020), and lack of recognizance (e.g., (Miles et al., 1974), what the literature lacks is a concrete analytical base on which managers comprehend why and when certain businesses do not identify and react to evidence of an environmental crisis, even though others may do. Whereas firms cannot react to every environmental change due to limited capital, it is critical that we consider the degree to which they can identify and react to different environmental forces. Perspectives on innovation center on the knowledge that is used to acquire new knowledge in an uncertain and context-dependent world. These features of innovation compel researchers to look for antecedents of innovation. The first antecedent of innovation is associated with innovation identification, which is the personal capacity to identify and eliminate uncertainty in information. The second antecedent of innovation is a crisis response, which is the firm reaction to significant external environmental change, especially amid the COVID-19 time.

We aim to propose new antecedents of these innovative actors driving the innovation as the dynamic capability to cope with challenges posed by the COVID-19 pandemic. The following research questions are proposed: Are innovation identification and pandemic responses enhancing a firm's innovation as an enact to the COVID-19 crisis's challenges and opportunities? What factors improve innovation identification and pandemic response during the transformation process? Understanding what factors affect innovation under the insight of knowledge-based perspective to address the recent and unforeseen factors that are antecedents to identify and respond to a new turbulent environment, and shape valuable solutions.

Specifically, we explore the data of 395 senior workers and managers in Vietnamese hospitality and tourism-related SMEs. This paper adds to the understanding and application of innovation as complex knowledge. To contribute to innovation studies, we propose the new notion of pandemic response and innovation identification as the major forces that empower the expansion of innovation in times of COVID-19 crisis. The concepts Covid-19 response and innovation-based identification reflect two aspects of knowledge-based innovation which are ambiguous and context-dependent. The relationships between pandemic response-innovation, and innovation identification-innovation may contribute to a sustainable innovation strategy.

LITERATURE REVIEW

Crisis-inspired Innovation

The global economy has been damaged by disasters. Firms have to face up with challenges of bankruptcy. Cutting down on innovation activities may put a firm's substance, key value-creating activities, knowledge base in danger (Zouaghi et al., 2018). However, firms who have a higher level of innovation-based identification would insist on their development path with innovative activities. Crises, including the Covid-19 pandemic we are currently facing, take a substantial economic and social impact, leaving properties and people stranded and creating significant social and economic disruption. Many dynamics, on the other hand, are ingredients for transformation, which contributes to the creation of new market models. The sharing digital economy, for example, emerged from the financial crisis of 2009, when technology allowed the development of marketplaces for underutilized assets at a time and people were actively finding alternative sources of income. Instantaneously, developments in digital technologies occur rapidly, including a digital platform, cloud-based computing, artificial intelligence (AI), big data analytics, Internet of Things (IoT), robotic systems, smart technologies, and 3-D printing. These technologies enormously transform an organization's functionality, human life, and addressing social business problems. During Covid-19, many advanced digital technologies are functioned to scanning infectious cases, contact tracing, and treating patients. During the quarantine, lockdown, and isolation, many firms adopt a digital platform to quickly restructure supply chains, collaborate with new partnerships, to control telework and distance education which will become the new innovative business model in the future.

Crisis advances the new technology which requires firms' behavioral response to technology to figures out the path to normal to reform economic and social structures (Sneader and Singhal, 2020). Technology helps to solve problems from treatment capacity of preventing the virus scattering, collaborate with new partners and set up new supply chain key, prevent material shortages (Bello et al., 2020). Firms' behavioral response quickly to pandemic could prompt urgency for innovation. Organizations may develop dynamic competencies based on speediness, versatility, and resilience to compete effectively in hyper-competitive and turbulent environments (Aghina, 2018). As a result, as an industry, government, or nonprofit organization, sustainable innovation has become a strategic priority (Veronica et al., 2020).

Covid-19 Sensemaking and Innovation-based Identification

A disaster creates an immediate and palpable sense of urgency. This sense of urgency allows organizations to put aside all other concerns and concentrate solely on a particular problem, reallocating resources if required. With a single emphasis and reallocated money, it is now everyone's responsibility to work together to address the crisis, adding a new diversity of views and experiences to the table. This sense of urgency and single-mindedness legitimizes what would otherwise be considered "waste," allowing for more innovation and learning. Since the situation is only immediate, the organization should devote itself to a concentrated innovation effort, access and interpret data from a variety of fields, resulting in a better understanding of future opportunities in a limited time (Roberts et al., 2016, Johnson and Murray, 2020).

We suggest how firms manage themselves and shed valuable insights on the way they organize to gain a strategic edge by out-of-the-box thinking. This line of investigation is consistent with previous research. Examining the contribution of information management using a knowledge-based view method and the effect of digital innovative solutions on the success of smaller businesses.

HYPOTHESES

This section sets out ten hypotheses that serve as the foundation for a knowledge transformation model. The critical elements of organizational innovation are identified as knowledge transfer in previous studies (Williams, 2007), as well as how they point to the need to respond to context and change knowledge through the transforming process. The concept of identification and response would concurrently operate in the knowledge transfer process. Two contingencies are suggested to result in higher identification and response stages, respectively. Finally, the study proposes determinants of identification and response to Covid-19 which generate knowledge transfer, which in a sequence contributes to improved innovation performance at the recipient.

Two critical features of organizational knowledge emerge from these facets of knowledge: causal complexity and context-dependency (Williams, 2007). Since knowledge is reflected in the organization's repetitive activities, known as rituals, causal ambiguity (Lippman and Rumelt, 1982) occurs. Routines connect the activities of members of an organization who may not recognize or even be mindful of the actions of those in the chain. Individual employees do not fully comprehend the association between a firm's activities and consequences since these series are lengthy and incompletely tacit (Nelson and Winter, 1982). Firms need the identification of knowledge resource to handle their knowledge characteristics of complexity. The majority of scholars believe that organizational knowledge is found in the organization's actions. The companies gain new knowledge by doing (Nelson and Winter, 1982), replication and adaptation (Williams, 2007). Similarly, Spender (1996) emphasizes the inextricable connection between awareness and collaborative action in organizations. Knowledge is profoundly embedded in action (Nonaka, 1994). It is critical for firms to identify with the progress of transferring knowledge.

Context dependency, on the other hand, occurs when knowledge incorporates components of knowledge that differ across contexts, such as individuals, personal networks, or circumstances. Since component convergence is required for successful knowledge (Argote and Ingram, 2000), setting can induce knowledge to be enacted in a mechanism that produces efficient results by incorporating different external and internal components. Firms need an appropriate response to the ambiguous context to transfer knowledge. Some approaches to understanding concentrate on the convergence of various knowledge elements. Knowledge, according to Argote and Ingram (2000), is stored in three repositories inside the firm: members, resources, activities, environment, and systems that connect them. Firm efficiency is determined by the degree to which these networks are integrated with one another and with the external environment. Ambiguous context forces firm response to deal with changes and turbulence.

Crisis-based innovation under the aspect of knowledge transfer necessitates firms to overcome a variety of unpredictable connections, including the relationship between an action and its consequence, the association between an operation and the rest of the organization, and the connection between an action and the firm's atmosphere, since unclear causality (innovation identification) and context-dependence (Covid 19-response) are central aspects of organizational knowledge. Firms may adjust a repetition, observe results, observe interactions, transform. Firms will eventually adopt behaviors at the source and combine them with the current environment to implement reaction and innovation. Thus, we propose the followings:

Hypothesis 1: Innovation-based identification with firms has a positive effect on the crisis-inspised innovation.

Hypothesis 2: The response to Covid-19 has a positive effect on the crisis-inspised innovation

Hypothesis 3: The response to Covid-19 has a positive effect on the innovation-based identification

Cooperative Culture

While innovation is goal-oriented and power-seekers crave status and credibility at work in order to achieve gratification by exercising their authority, and they have the proclivity to follow their own individualistic goals, such as achieving managerial and political roles (Anderson et al., 2014). Finally, the need to response to climate change, form warm and cooperative relationships with others is referred to as the need for association (Sánchez García and Díez Sanz, 2018). During the pandemic-induced organizational crisis, individuals who have clear association motivations in their relationships are more cooperative and beneficial. People with strong affinity motivations perform best in teams made up of cooperative people, according to research, and a collectivistic community arises that encourages mutual interests and

cooperative activities while putting a premium on optimizing collective wellbeing (Nan and Lu, 2014). Although the notion that innovative corporate cultures can increase internal motivation has a long tradition (e.g., Hackman and Oldham 1980), it has gotten little scientific recognition (Lee et al. 2004). According to reviews of innovation studies, innovative workplace culture can be a major factor in deciding how much innate motivation workers have (Saldanha et al., 2021). When employees believe that the corporate atmosphere promotes coworker teamwork in achieving organizational objectives, it is anticipated that there is a clear positive association between a strong goal-setting and accomplishment of innovation. As a consequence, we suggest the following:
Hypothesis 4: Corporative culture has a positive effect on the individual's innovation-based identification of firms.

Innovative culture

A risk-taking, results-oriented, exciting, demanding, and enterprising job climate is what a creative community means in the broadest context. Employees are expected to be innovative and dynamic in terms of inspiration, breadth, depth, and cultural distance (Godart et al., 2014). This culture, which emphasizes novelty, equity among participants, transparency, and versatility, is thought to encourage creativity by sending out powerful signals about norms and values that mean it is healthy for individuals to engage in risky, exploratory, and failure-prone behaviors associated with creativity. The previous study has discovered a positive correlation between high high-involvement work system and creativity (Li et al., 2017). Since high achievers strive to be the best and are willing to take chances, they can thrive in a creative environment (Snihur and Zott, 2019). Individuals with a high-involvement work system, cultural dimensions related to goal-setting and accomplishment (e.g., individualism, masculinity, and long-term orientation) are thus expected to do well in a creative society that allows them to follow high organizational expectations (Saldanha et al., 2021). As a consequence, we suggest the following hypothesis:
Hypothesis 5: Innovative culture has a positive effect on innovation-based identification.
As business strategies adjust to new consumer realities, strategic advantages change dynamically, and the essential skills that once differentiated an enterprise. The ongoing crisis has greatly amplified and escalated their destructive power. Modifications of distribution templates. To compete, companies with large ground forces do not depend on individuals. Firms have adapted and changed their go-to-market approach since the pandemic struck, with the vast majority focusing on various ways of digital interaction with consumers. As marketers learn that augmented reality enables them to do tasks that were previously virtually impossible, such as creating the "right squad" of specialists for any promotional presentation, sales coverage has been fully redefined. Smaller companies will also "match up" to their bigger counterparts in the digital distribution sphere (Sneader and Singhal, 2020). Understanding those shifts and the challenges they create will help businesses achieve long-term advantages. During previous crises, businesses that invested in novelty saw superior post-crisis development and profitability. Organizations that stayed focused on creativity through the financial crisis of 2009, for example, appeared bigger, outpacing the economy by more than 30% and remaining to expand at a faster pace for the next three to five years. Thus, we propose the following hypothesis:
Hypothesis 6: Adaptation has a positive effect on innovation-based identification.
The pandemic response of the firm can be defined as the propensity to embrace and use new technologies for accomplishing goals (readiness), activities aimed at entering the new product and process domains (exploration), activities aimed at improving the existing product and process positions (exploitation) (Lee and Grewal, 2004, Jafari-Sadeghi, 2021). This review

seeks to capture the scope of digital technologies for the organizational response to COVID-19 around the world, as well as their shortcomings and implementation challenges, such as legal, ethical, and privacy issues, as well as organizational and staff issues (Budd et al., 2020). Based on the above arguments, we suppose that a firm with a large magnitude can result in high response (Zouaghi et al., 2018).

Hypothesis 7: Magnitude has a positive effect on the pandemic response of firms.

When it comes to a firm's controllability, it's necessary to distinguish between the manageablility of the phenomenon itself (the source) and its effect. This is because the firm's significant decision is influenced by a conviction in the degree to which the company can manipulate and interact with the source and/or the outcomes of it (Plambeck and Weber, 2009). The controllability of a company is based on its subjective conviction of self-efficacy. A company may face disasters due to a lack of responses (Ashforth et al., 2014). The problem is that in such a situation the lack of response is not due to the lack of data/information by the organization. Thus, we propose:

Hypothesis 8: Controllability has a positive effect on the pandemic response of firms.

Status-certainty as well as environmental certainty and responses status-certainty/uncertainty has been regarded as a crucial determinant of management activity (Ashill and Jobber, 2009). Since detecting environmental risks and opportunities takes a decision (Milliken, 1987), there is variation in how managers react to identical environmental factors (Milliken, 1990) or in how they respond (Grandori, 1984). Companies are likely to vary in both their particular responses to external events and the outcomes of such responses (Lee and Grewal, 2004). In general, a high degree of status-certainty assurance boosts managers' trust in their knowledge of the market climate, making it easier for them to determine whether or not to react to the pandemic situations.

Hypothesis 9: Status-certainty has a positive effect on the pandemic response of firms

The operational firms in markets characterized by the rapid technological revolution and quick product life cycles are forced to constantly incorporate new technological technologies and inventions to compete effectively (Nieves and C., 2015). So that tech-speed is expected to have positive relationship with the crisis response of a firm.

Hypothesis 10: Tech-speed has a positive effect on the pandemic response of firms

METHODS

Vietnamese hospitality and tourism-related firms were chosen as the sample for this study. There are reasons for this selection: the pandemic response and innovation-based identification amid the Covid-19. About the fact that Vietnam is a heavily centralized nation, a lot of main decisions were taken at the local level, which helped the country respond quickly (OurWorldinData, 2021). While tourism is the most vulnerable to the pandemic, firms have transformed to be digital innovators for travel-related bookings has been increasing even before the pandemic. Travelers all over the world are customizing their vacations with destination journeys. Spending on tourism is moving away from hotels and toward travelers' experience, which is an innovation progress identification in Vietnam (Constantin et al., 2021).

The data has been collected by the Small and Medium Enterprise Association, who possesses an available list of 1000 SME firms in hospitality and tourism. They are employed to perform the survey from May 2020 to December 2020. After data cleaning, there were 395 usable responses for the analysis.

RESULTS

All validity and reliability have been statistically assessed. All the hypothesised are supported. The details are reported as following tables:

Table 1: Description, Correlation, Reliability, and Discriminant Validity									
	IN	IC	AD	CO	SC	TS	RS	ID	CI
IN									
IC	.471**								
AD	.345**	.400**							
C	.368**	.544**	.324**						
SC	.387**	.535**	.348**	.516**					
TS	.405**	.528**	.399**	.444**	.455**				
RS	.360**	.478**	.366**	.545**	.473**	.391**			
ID	.414**	.491**	.388**	.488**	.465**	.430**	.426**		
IC	.469**	.549**	.483**	.472**	.526**	.464**	.506**	.535**	

*Note: *p < 0.05, ** p < 0.01, IN = innovativeness, IC = Innovative Collaboration, AD= Adaptation, CO= Controlability, SC = status certainty, TE = technology Speed, RE= COVID-19 response, ID=Identificication, OI= crisis-inspired innovation, SD = Standard deviation, CR = Composite reliability. The square root of the AVE is in parentheses. The numbers in parentheses are all higher than the corresponding off-diagonal elements.*

Table 2. Descriptive Statistics and PLS-CFA Results (n=395)									
	FL	Mean	SD	T Value	P Values	VIF	Cronbach's Alpha	Composite Reliability	Average Variance Extracted (AVE)
AdInt1	0.831	5.329	1.255	55.753	0.000	1.852	0.827	0.827	0.827
AdInt2	0.788	5.42	1.202	44.235	0.000	1.679			
AdInt3	0.819	5.401	1.241	50.298	0.000	1.811			
AdInt4	0.807	5.486	1.206	46.186	0.000	1.751			
CoCul1	0.842	5.543	1.144	67.615	0.000	1.985	0.810	0.875	0.637
Cocul2	0.744	5.537	1.134	32.646	0.000	1.520			
Cocul3	0.832	5.589	1.15	62.196	0.000	1.932			
Cocul4	0.771	5.568	1.138	35.717	0.000	1.615			
Con1		5.537	1.153			1.000		1.000	1.000
IIden1	0.849	5.567	1.115	69.668	0.000	1.653	0.736	0.850	0.655
IIden2	0.784	5.503	1.169	45.844	0.000	1.392			
IIden3	0.793	5.528	1.179	35.123	0.000	1.446			
InCul1	0.823	5.525	1.211	45.634	0.000	1.883	0.838	0.891	0.673
InCul2	0.826	5.591	1.182	54.755	0.000	1.933			
InCul3	0.801	5.576	1.156	45.870	0.000	1.742			
InCul4	0.830	5.587	1.188	61.794	0.000	1.999			
Mgni1	0.849	5.488	1.144	72.087	0.000	1.633	0.740	0.853	0.659
Mgni2	0.763	5.495	1.227	37.887	0.000	1.343			
Mgni3	0.821	5.505	1.154	54.073	0.000	1.555			
OrInn1	0.825	5.49	1.18	65.154	0.000	1.801	0.806	0.873	0.632

OrInn2	0.783	5.58	1.118	44.588	0.000	1.635			
OrInn3	0.787	5.539	1.104	42.798	0.000	1.667			
OrInn4	0.784	5.532	1.16	48.769	0.000	1.664			
Res1	0.807	5.554	1.153	56.508	0.000	1.651	0.785	0.861	0.608
Res2	0.783	5.528	1.131	43.145	0.000	1.592			
Res3	0.760	5.55	1.128	42.115	0.000	1.505			
Res4	0.767	5.442	1.223	35.122	0.000	1.536			
StCer1	0.834	5.549	1.113	53.672	0.000	1.591	0.745	0.855	0.662
StCer2	0.789	5.527	1.161	40.260	0.000	1.371			
StCer3	0.819	5.479	1.16	45.546	0.000	1.566			
TeS1	0.807	5.501	1.144	53.312	0.000	1.857	0.829	0.880	0.595
TeS2	0.742	5.517	1.186	33.553	0.000	1.581			
TeS3	0.777	5.514	1.145	42.037	0.000	1.818			
TeS4	0.730	5.413	1.254	30.480	0.000	1.520			
TeS5	0.797	5.463	1.115	49.456	0.000	1.876			

We examined framework as these hypotheses propose a direct and positive effect of independent variables on dependent variables.

Table 4. PLS Coefficient Path Analysis

Path	Mean	Standard Deviation	T Statistics	P Values	VIF
H1: Identification -> Innovation	0.454	0.045	10.076	0.000	1.762
H2: Response -> Innovation	0.460	0.043	10.730	0.000	1.762
H3: Response -> Identification	0.328	0.056	5.850	0.000	2.068
H4:Corporative culture -> Identification	0.245	0.053	4.576	0.000	2.911
H5:Innovative culture -> Identification	0.174	0.047	3.690	0.000	2.335
H6: Adaptation -> Identification	0.164	0.050	3.265	0.001	1.612
H7: Magnititude -> Response	0.421	0.054	7.754	0.000	2.440
H8: Controllability -> Response	0.128	0.034	3.752	0.000	1.743
H9: Status_certainty -> Response	0.152	0.048	3.185	0.002	2.460
H10: Tech_speed -> Response	0.279	0.048	5.866	0.000	2.201

DISCUSSION AND CONCLUSIONS

We live in a digital era marked by dynamic, chaotic, and huge changes. The creative world and digital transformation have resulted from the compounding impact of multiple market forces. Organizations must be flexible with dynamic skills to survive and succeed in this modern business setting. Market gyrations (e.g., economic downturns), political instablities (e.g., Myanmar coup geopolitics and US-China trade wars), climate change, wars, and disease problems (e.g., the global pandemic of COVID-19) all put firms to the eventual test of their long-term viability. Organizations must depend on their dynamic capabilities to not only survive but also thrive in times of crisis. Enterprises, states, and nonprofits have all realized the importance of long-term innovation.

On the basis of this research, a number of practical implications can be made. First, the major finding concerns the firm's innovative culture, corporative culture, adaptation being significantly influence on innovation-based identification which leads to crisis-inspired innovation.

The tourism firms have innovatively transformed their strategic focus to the domestic market, which it did not receive the recognition it deserved before the pandemic. The demand, tastes, patterns, and favorite products of visitors, as well as tourism services that serve citizens have not well defined. And after the pandemic is fully under control, the related agencies should establish a robust domestic travel industry. This competition would ensure the long-term growth of the whole industry if adjustment strategies are applied. Tourism growth strategies in Vietnam must prioritize protection, safety, and environmental values such as peace with nature, respect for local culture and identity, positive disease response, and long-term sustainability in order to promote domestic tourism. In order to satisfy the tourists and travellers' expectations, innovative solutions in terms of safety and sanitary measures, as well as the closeness of medical facilities, will be critical.

The collaborative efforts of leaders from travel management agencies and enterprises around the country to restore tourism in the new normal state. Both tourism service providers, management agencies, and related ministries will work together to build a supportive environment for boosting domestic tourism production, ensuring the tourism sector's activity and growth in the new normal state.

The innovative culture of tourism may link and diversify products. Travel companies should restructure and recalculate relevant and crucial facets of their domestic travel operations. Many visitors book experiences in advance of their trip, implying that the in-destination experience has a greater influence on the overall traveler decision-making development. Many outdoor trips have been the key incentive for visitors to visit a destination in the first place, such as cave exploration, highland camping, remote island visits, sport activities, and food and beverage festivals (Constantin et al., 2021).

Technology speed increase the firm's response to the pandemic. Local operators, who also lack the expertise of larger travel agencies but are more flexible when it comes to planning customized events may use increasingly common online players to communicate directly with consumers and offer these options. Companies in the tourism industry may decide to divert their struggles breakthrough from developing resorts and vending exploration tickets to creating exclusive events and using these channels to capitalize on travel-experience patterns (Constantin et al., 2021). Tourism industry will undoubtedly rely heavily on digital and technological breakthroughs such as the use of virtual agents and chatbots to make reservations, mobility patterns to control visitor flows, service-oriented robotics, artificial intelligence, the Internet of Things, 5G, and gamification methods for emotional management. The future visitor experience will be shaped by these technological advances.

At its most basic level, continued environmental scanning is essential for increasing state certainty. Once state certainty has been increased, response paths direct become significant. This is particularly true in Vietnam, where the majority of visitors come from neighboring countries with close economic links and low transmission rates. Travel companies could benefit from the environmental state certainty to response and monitor the growth of travel bubbles to catch early outbound demand.

Since the Vietnamese government can control the pandemic, it is critical for an organization to achieve a greater sense of controllability if it is more of a "discovering" type of organization than an "enacting" type of organization. Firm members of the company must consider what is at stake for the organization as a result of the pandemic crisis. Finally, a matrix-based organizing structure can be given to managers who need to explore and conceptualize a wide variety of possible responses to the crisis in terms of magnitude and scope.

The study highlights a wide range of under-researched study opportunities on firms' responses to external factors that are relevant to but not explicitly controllable by any particular entity. Natural disasters, for example, are sudden, measurable, well-known, scarcely manageable, low-probability, high-influence calamity. Many disasters (e.g., hurricanes, tornadoes, and earthquakes) provide specific knowledge that can help us predict and plan for them (e.g., that a specific type of disaster is more likely to happen sooner than later; that a certain area is more vulnerable to a specific type of disaster). However, it is difficult for an organization to make precise forecasts about when a catastrophe will occur (state uncertainty); what impact it will have on an organization; and, as the result, it is difficult for the firm to determine whether and in what way it should plan and react. In a decision-making state like this, cognitive Covid-19 should be prominent, and we need to learn more about the antecedents and implications. Finally, by examining the impact of Covid-19, we add to the body of knowledge-based view in the field of firm cognition. So far, there is no specific guidance on the effects of the relation between sensemaking and reponse in the literature. There may be a mediating and nonlinear relationship between managerial response, identification, and innovation.

Acknowledgement

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 502.02-2019.310

REFERENCES

Aghina, W., De Smet, A., Lackey, G., Lurie, M., and Muraka, M. 2018. The five trademarks of agile organizations. McKinsey & Company Report, January 22.

Am, J. B., Furstenthal, L., Jorge, F. & Roth, E. 2020. Innovation in a crisis: Why it is more critical than ever. McKinsey & Company, June 17.

Anderson, N., Potočnik, K. & Zhou, J. 2014. Innovation and Creativity in Organizations: A State-of-the-Science Review, Prospective Commentary, and Guiding Framework. Journal of Management, 40, 1297-1333.

Ashforth, B., Rogers, K., Pratt, M. & Pradies, C. 2014. Ambivalence in Organizations: A Multilevel Approach. Organization Science, 25, 1453-1478.

Ashill, N. J. & Jobber, D. 2009. Measuring State, Effect, and Response Uncertainty: Theoretical Construct Development and Empirical Validation. Journal of Management, 36, 1278-1308.

Bello, J., Collins, S., Dreischmeier, R. & Libarikian, A. 2020. Innovating from necessity: The business-building imperative in the current crisis. McKinsey Digital, April 16.

Brosseau, D., Ebrahim, S., Handscomb, C. & Thaker, S. 2019. The journey to an agile organization. McKinsey & Company, May 10.

Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., Rees, G., Emery, V. C., Stevens, M. M., Keegan, N., Short, M. J., Pillay, D., Manley, E., Cox, I. J., Heymann, D., Johnson, A. M. & Mckendry, R. A. 2020. Digital technologies in the public-health response to COVID-19. Nature Medicine, 26, 1183-1192.
Colombo, M. G., Piva, E., Quas, A. & Rossi-Lamastra, C. 2020. Dynamic capabilities and high-tech entrepreneurial ventures' performance in the aftermath of an environmental jolt. Long Range Planning, 102026.
Constantin, M., Francois, M. & Le, T. 2021. Reimagining tourism: How Vietnam Reimagining tourism: How Vietnam can accelerate travel recovery can accelerate travel recovery. McKinsey & Company, March 19.
Dunning, J. & Lundan, S. 2009. The institutional origins of dynamic capabilities in multinational enterprise. Industrial and Corporate Change, 19, 1225-1246.
Ebersberger, B. & Kuckertz, A. 2021. Hop to it! The impact of organization type on innovation response time to the COVID-19 crisis. Journal of Business Research, 124, 126-135.
Elenkov, D. S. 1997. STRATEGIC UNCERTAINTY AND ENVIRONMENTAL SCANNING: THE CASE FOR INSTITUTIONAL INFLUENCES ON SCANNING BEHAVIOR. Strategic Management Journal, 18, 287-302.
Fainshmidt, S., Pezeshkan, A., Lance Frazier, M., Nair, A. & Markowski, E. 2016. Dynamic Capabilities and Organizational Performance: A Meta-Analytic Evaluation and Extension. Journal of Management Studies, 53, 1348-1380.
Godart, F. C., Maddux, W. W., Shipilov, A. V. & Galinsky, A. D. 2014. Fashion with a Foreign Flair: Professional Experiences Abroad Facilitate the Creative Innovations of Organizations. Academy of Management Journal, 58, 195-220.
Johnson, E. & Murray, F. 2020. What a Crisis Teaches Us About Innovation. MIT Sloan Management Review, November 30.
Karna, A., Richter, A. & Riesenkampff, E. 2016. Revisiting the role of the environment in the capabilities–financial performance relationship: A meta-analysis. Strategic Management Journal, 37, 1154-1173.
Kraus, S., Clauss, T., Breier, M., Gast, J., Zardini, A. & Tiberius, V. 2020. The economics of COVID-19: initial empirical evidence on how family firms in five European countries cope with the corona crisis. International Journal of Entrepreneurial Behavior & Research, 26, 1067-1092.
Lee, R. P. & Grewal, R. 2004. Strategic Responses to New Technologies and Their Impact on Firm Performance. Journal of Marketing, 68, 157-171.
Lejarraga, J. & Pindard-Lejarraga, M. 2020. Bounded Rationality: Cognitive Limitations or Adaptation to the Environment? The Implications of Ecological Rationality for Management Learning. Academy of Management Learning & Education, 19, 289-306.
Li, Y., Wang, M., Van Jaarsveld, D. D., Lee, G. K. & Ma, D. G. 2017. From Employee-experienced High-involvement Work System to Innovation: An Emergence-based Human Resource Management Framework. Academy of Management Journal, 61, 2000-2019.
Miles, R., Snow, C. & Pfeffer, J. 1974. Organization-Environment: Concepts and Issues. Industrial Relations: A Journal of Economy and Society, 13, 244-264.
Nan, N. & Lu, Y. 2014. Harnessing the Power of Self-Organization in an Online Community During Organizational Crisis. MIS Quarterly, 38, 1135-1158.
Nieves, J. & C., M. S. 2015. Management innovation in the hotel industry. Tourism Management, 46, 51-58.
Ourworldindata 2021. Emerging COVID-19 success story: Vietnam's commitment to containment. OurWorldinData.org, June 30, 2020, Retrieved from https://ourworldindata.org/covid-exemplar-vietnam-2020.
Plambeck, N. & Weber, K. 2009. CEO Ambivalence and Responses to Strategic Issues. Organization Science, 20, 993-1010.

Roberts, N., Campbell, D. E. & Vijayasarathy, L. R. 2016. Using Information Systems to Sense Opportunities for Innovation: Integrating Postadoptive Use Behaviors with the Dynamic Managerial Capability Perspective. Journal of Management Information Systems, 33, 45-69.
Saldanha, T. J. V., John-Mariadoss, B., Wu, M. X. & Mithas, S. 2021. How Information and Communication Technology Shapes the Influence of Culture on Innovation: A Country-level Analysis. Journal of Management Information Systems, 38, 108-139.
Sánchez García, J. L. & Díez Sanz, J. M. 2018. Climate change, ethics and sustainability: An innovative approach. Journal of Innovation & Knowledge, 3, 70-75.
Sneader, K. & Singhal, S. 2020. Beyond the coronavirus: The path to the next normal. McKinsey & Company, March 23.
Snihur, Y. & Zott, C. 2019. The Genesis and Metamorphosis of Novelty Imprints: How Business Model Innovation Emerges in Young Ventures. Academy of Management Journal, 63, 554-583.
Veronica, S., Alexeis, G.-P., Valentina, C. & Elisa, G. 2020. Do stakeholder capabilities promote sustainable business innovation in small and medium-sized enterprises? Evidence from Italy. Journal of Business Research, 119, 131-141.
Williams, C. 2007. Transfer in context: Replication and Adaptation in knowledge transfer relationships. Strategic Management Journal, 28, 867-889.
Zouaghi, F., Sánchez, M. & Martínez, M. G. 2018. Did the global financial crisis impact firms' innovation performance? The role of internal and external knowledge capabilities in high and low tech industries. Technological Forecasting and Social Change, 132, 92-104.

DECISION SCIENCES INSTITUTE

Dimensions, Management and Outcomes of Supply Base Complexity: A Systematic Literature Review

Melek Akın Ateş
Sabanci Business School, Sabanci University
Email: melek.ates@sabanciuniv.edu

Davide Luzzini
EADA Business School
Email: dluzzini@eada.edu

ABSTRACT

This research aims to propose a comprehensive conceptualization of supply base complexity (SBC), evaluate its performance implications, and provide avenues for future research. For this purpose, a systematic literature review is conducted, resulting in a final set of 91 articles across 26 journals published in the last two decades (2002-2021). We distinguish three SBC dimensions (i.e. structural, operational, and relational) and several sub-dimensions. Furthermore, we provide an overview of the nuanced picture of the potential SBC effects on several performance dimensions, by also illustrating how SBC can be managed. Finally, we pose research questions that can guide future research on SBC.

KEYWORDS: Systematic literature review, Supply base, Complexity, Performance

INTRODUCTION

Purchasing and supply management (PSM) essentially entails the “strategic approach to planning for and acquiring the organization's current and future needs through effectively managing the supply base” (Monczka et al., 2015). Central to PSM is the idea that a focal firm is able to successfully manage its supply base - suppliers that are actively managed through contracts and the purchase of parts, materials and services (Choi and Krause, 2006, p. 639) - in a way that is strategically aligned with the firm’s business strategy (González-Benito, 2007), ultimately generating a sustainable competitive advantage. However, supply bases are getting increasingly complex and organizations increasingly have more difficulty to successfully manage their supply base (Bode and Wagner, 2017; Lu and Shang, 2017).

Choi and Krause (2006) lay the foundations of supply base complexity (SBC), defining it as “the degree of differentiation of the focal firm’s suppliers, their overall number, and the degree to which they interrelate” (p. 638). SBC is often associated with detrimental outcomes such as increased supply disruptions, higher costs and delivery problems (Bozarth et al., 2009; Brandon-Jones et al., 2015). Large supply bases consisting of many suppliers with high volatility and uncertainty increase coordination costs and reduce visibility (Bode and Wagner, 2015; Brandon-Jones et al., 2014; Choi and Krause, 2006). However, other scholars perceive SBC as a strategic asset that can be utilised to increase the likelihood of generating innovations (Choi and Krause, 2006; Lu and Shang, 2017) and provide firms with the ability to recover from disruptions (Birkie and Trucco, 2020). While the importance of SBC is acknowledged by scholars and practitioners, the mixed findings make it difficult to come up with a panacea and identify complexity management practices (Aitken et al., 2016; Turner et al., 2018).

Currently, the research landscape around SBC is scattered around several overlapping and/or complementary sub-dimensions of SBC and there is a mixed evidence on its outcomes. Some dimensions clearly emerge since the foundational literature, such as the structural characteristics of a system (i.e. number and qualities of elements) or the interactions between the elements (Thompson, 1967). Others are more blurry defined, such as variety, diversity, variation, and uncertainty (Isik, 2010). In the end, although several disciplines recognize the important role of SBC, it remains an “elusive” construct (Bode and Wagner, 2015).
In this study, we aim to provide a comprehensive review of the previous studies and present an integrative view of SBC, by not only clarifying how SBC is conceptualized, but also assessing its performance implications as well as complexity management practices. In line with this objective, we aim to answer the following research questions:

RQ1: What does SBC mean? i.e. What are the key dimensions of SBC?
RQ2: What is the impact of SBC on firm performance?
RQ3: How can firms manage SBC?

The remainder of this paper is structured as follows. First, we provide a background of SBC by investigating the existing perspectives and theories. Then, we discuss our research method and state our database search procedure. Afterwards, we present the results of systematically analyzing 91 relevant studies in four main parts: *i) descriptive results, ii) SBC dimensions, iii) SBC outcomes,* and *iv) managing SBC*, and summarize the key variables and dimensions in an integrated framework. Finally, we discuss the results, make suggestions for future research and present the conclusions and limitations.

BACKGROUND: EXISTING PERSPECTIVES ON SUPPLY BASE COMPLEXITY

One of the prevailing perspectives that can inform the study of SBC is the *entropy-based approach,* which is based on Shannon’s (1948) information entropy. Some scholars apply this perspective to the supply chain as the unit of analysis (e.g. De Leeuw et al. 2013; Dittfeld et al., 2018; Frizelle and Woodcock, 1995; Isik, 2010; Serdarasan, 2013); considering not only suppliers, but also customers and internal company operations, as well as materials and information flows across the supply chain. Nevertheless, we believe that some relevant concepts can be carried into the context of SBC, while others are out of scope. In particular, supply chain complexity is classified into *structural* (or static, also referred to as detail) and *operational* (or dynamic) complexity. Structural complexity deals with variety and operational complexity deals with uncertainty.

The entropy-based approach is a cornerstone of the complexity debate and still represents the anchor of numerous supply chain studies. A close and consistent research stream in social sciences is the *complex systems perspective,* which is usually traced back to Simon (1962, p. 468) who stated that a socio-technical system is complex if it is “made up of a large number of parts that interact in a non-simple way.” This definition has been the basis for many subsequent conceptualizations of complexity and allows to further clarify the distinction between structural and operational complexity. The former refers to the number and variety of elements defining the system; while the latter refers to unpredictability, randomness, or frequent changes in a system's response to a given set of inputs (Bozarth et al., 2009; Bode and Wagner, 2015). As Bode and Wagner (2015) observe, “these aspects are often closely interrelated, because the larger the number of varied elements, the greater the possible number of interactions and thus the variety of behaviors and states the system may exhibit. This is especially true of supply chains."

This view is reinforced by a lateral stream of research that we can label as the *social capital* perspective. We do not aim entering the details of social capital theory or the definitions of social capital, but the reader might refer for example to Nahapiet and Ghoshal (1998). Yet, it is worth noting that social capital is usually split into different components that seem consistent with the previously mentioned perspectives. Indeed, some supply chain scholars have investigated the formation of social capital in the context of buyer-supplier relationships (e.g. Carey et al., 2011; Horn et al., 2014; Roden and Lawson, 2014) or supply network (e.g. Choi and Kim, 2008; Hartmann and Herb, 2015; Kim, 2014; Longoni and Luzzini, 2016), leading to a distinction of a structural and relational component of social capital. For instance, Kim (2014) refers to structural embeddedness as the position in an extended network; while relational embeddedness refers to the ties, the relational assets, and the strength of interactions in the network.

Finally, we can mention a few studies that do not link to a specific theory, but address the topic of complexity at the manufacturing level. An example is the study of Jacobs and Swink (2011), who define product portfolio architectural complexity as a design state manifested by the multiplicity, diversity, and interrelatedness of products within the portfolio. Later, Jacobs (2013) also proposed a generalization and empirical example of such complexity measure. Again, we can observe that concepts like multiplicity, diversity and interrelatedness are consistent with the dimensions of complexity introduced so far.

As a result of this preliminary overview, we could find a summary of relevant studies about complexity in the supply chain domain, even though a systematic review of the literature is not available. Moreover, it is only with the *complex adaptive system (CAS) perspective* that scholars started explicitly addressing SBC (Choi and Krause, 2006). Choi et al. (2001) recognize supply networks as complex adaptive systems and suppliers as autonomous actors. Taking the perspective of the focal company, Choi and Krause (2006) conceive SBC as "a factor of the number of suppliers in the supply base, the level of supplier interaction, and the degree to which these suppliers vary in terms of organizational culture, size, location, technology, and so on." Accordingly, SBC is defined in terms of number, differentiation, and interrelationships among suppliers.

Since Choi and Krause's is probably the first study conceptualizing SBC, we believe it is time to re-examine state-of-the-art of the literature. In order to provide an integrative view and hopefully lead to the emergence of a dominant definition of SBC and its implications, we conducted a systematic literature review, as explained hereafter.

METHODOLOGY

To provide a comprehensive conceptualization of SBC and assess it performance implications, we conducted a systematic literature review via a detailed database search. We focused on full-text, peer-reviewed and scholarly articles in English, and conducted our literature search in the ABI/INFORM Global (Proquest) database. Our focus is on complexity at the upstream supply chain, i.e. SBC; however, SBC is often discussed as a sub-dimension of supply chain/network complexity. Furthermore, some studies also use slightly different terminology such as "supply chain structure/design". To account for this variety and not miss out any crucial studies, we used the following search string to identify relevant articles: *"Supply complexity"* OR *"Supply structure"* OR *"Supply design"* OR *"Supply base complexity"* OR *"Supply base structure"* OR *"Supply base design"* OR *"Supply chain complexity"* OR *"Supply chain structure"* OR *"Supply chain design"* OR *"Supply network complexity"* OR *"Supply network structure"* OR *"Supply network design"*. Following the recent guidelines of Durach et al. (2017) about conducting systematic literature reviews in SCM, we included empirical studies (quantitative and qualitative), conceptual studies, and modelling and simulation studies. The initial list consisted of 4017 articles. To assess relevance, we first read the abstracts of the articles. If SBC was not mentioned in the abstract, we further checked the full paper in order to not miss any relevant articles. If performance implications were not investigated, but nonetheless SBC dimensions were discussed, we still included them in the review. We excluded several articles which mentioned SBC only a few times, without being one of the key constructs in the study. After these steps, our final sample consisted of 91 articles.

To discuss descriptive characteristics and explain the differences in study findings (Durach et al., 2017), we first coded the articles based on the study context (i.e. country and industry) and research methods. Our main coding effort focused on identifying dimensions of SBC and the impact on firm performance. We mainly adopted and combined the existing classifications of SBC (e.g. Bode and Wagner, 2015; Bozarth et al., 2009; Brandon-Jones et al., 2015; Choi and Krause, 2006; De Leeuw et al., 2013). One of the main classifications focuses on the level of analysis in the supply chain by distinguishing between internal, downstream, and upstream complexity. In this systematic literature review, we only focus on upstream (supply base) complexity. We then identified three main categories of SBC: *structural, operational,* and *relational*. All main categories were further divided into sub-categories which are discussed in detail in the next sections. In order to assess the implications of SBC, we examine three key dimensions of firm performance*: operational, market*, and *financial*. This classification has also been frequently used in supply chain research, and specifically in literature reviews and meta-analyses (e.g. Golicic and Smith, 2013; Hult et al., 2008). Finally, we also discuss a recent emerging research stream, i.e. ‘managing’ SBC.

RESULTS: DESCRIPTIVE FINDINGS

We first present the descriptive findings in terms of i) time period, ii) publication outlet, iii) industry and country, and iv) research method.

Time Period

We find that out of 91 studies in our sample, 57% of the articles have been published in the last decade (see Figure 1). It is interesting to note that in 2015 there was a peak with 11 studies, and since then there has been a slight decreasing trend.

Publication Outlet

The 91 papers included in the literature review were published in 26 different journals (see Table 1). Approximately half of the studies in the sample were published in four journals: *International Journal of Operations & Production Management, International Journal of Physical Distribution & Logistics Management, Journal of Operations Management,* and *Journal of Supply Chain Management*.

Figure 1: SBC publications by year (2002-2021)

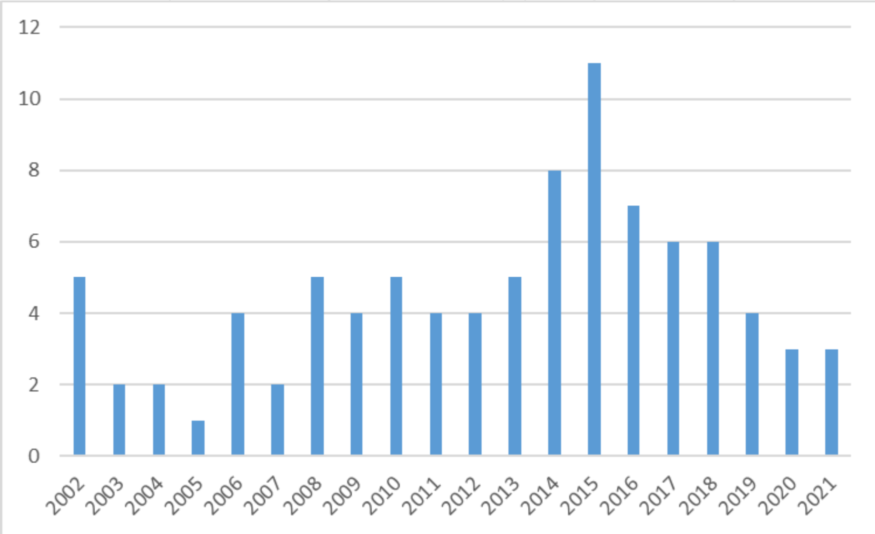


Table 1: SBC publications per journal

Journals	Fre.	%
Academy of Management Journal	1	1.1%
Applied Mathematical Modelling	1	1.1%
Computers & Industrial Engineering	2	2.2%
Decision Sciences	6	6.6%
European Journal of Operational Research	2	2.2%
IEEE Transactions on Engineering Management	1	1.1%
Information Systems Journal	1	1.1%
International Business Review	1	1.1%
International Journal of Logistics Research and Applications	2	2.2%
International Journal of Operations & Production Management	10	11.0%
International Journal of Physical Distribution & Logistics Management	9	9.9%
International Journal of Production Economics	5	5.5%
International Journal of Production Research	4	4.4%
Journal of Business Logistics	3	3.3%
Journal of Management Information Systems	1	1.1%
Journal of Manufacturing Technology Management	1	1.1%
Journal of Operations Management	13	14.3%
Journal of Purchasing and Supply Management	3	3.3%
Journal of Supply Chain Management	8	8.8%
Management Decision	1	1.1%
Omega	1	1.1%
Production Planning & Control	1	1.1%
Resources Policy	1	1.1%
Supply Chain Forum: International Journal	2	2.2%
Supply Chain Management: An International Journal	6	6.6%
The International Journal of Logistics Management	5	5.5%
Total	*91*	*100%*

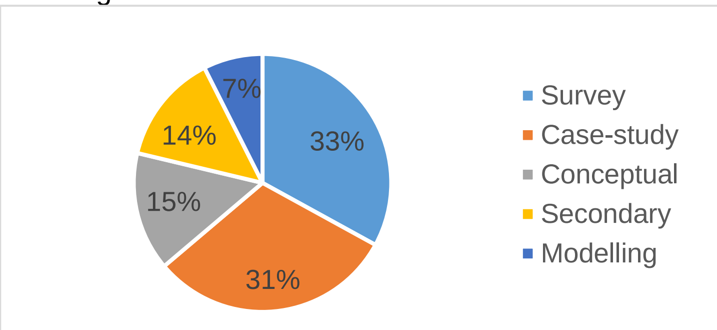
Industry and Country

Excluding articles with conceptual and modelling approaches, we examined the countries and industries of the data sets in articles. We found that out of the 68 articles, 62% represent single-country and 21% represent multi-country data collection efforts (in 12 articles, countries were not explicitly mentioned). USA appears the most (28) in both single and multi-country studies, followed by Germany (10) and the Netherlands (11), and Spain (9). Regarding the industries, 65% are multi-industry and 32% are single-industry studies (In three studies industry was not explicitly mentioned).

Research Method

The dominant research methods in SBC research seem to be surveys (33%) and case studies (31%). The next popular research strategy is conceptual studies (15%); whereas some of these studies focus on conceptualising SBC (i.e. Choi and Krause, 2006; Manuj and Sahin; 2011; Serdarasan, 2013) others only partly discuss SBC and in relation to other concepts, rather than as the main concept (i.e. Defee and Stank, 2005; Lambert and Cooper, 2000; Manuj and Mentzer, 2008). Several of these conceptual studies also aim to examine performance implications such as cost, risk, innovativeness, sustainability, and transparency (e.g. Autry and Griffis; 2008; Choi and Krause, 2006; Skilton and Robinson, 2009; Tachizawa and Wong, 2015). We see that 14% of the papers use secondary data, especially when a “network” perspective is adopted (e.g. Sharma et al., 2019). The first study using secondary data was published in 2013, but it gained popularity increasing its share to 32% of the total studies published in the last five years. Interestingly, none of the studies adopts experiments as the research method. Finally, as we did not limit our systematic literature review to only empirical research methods, modelling papers are also represented (7%).

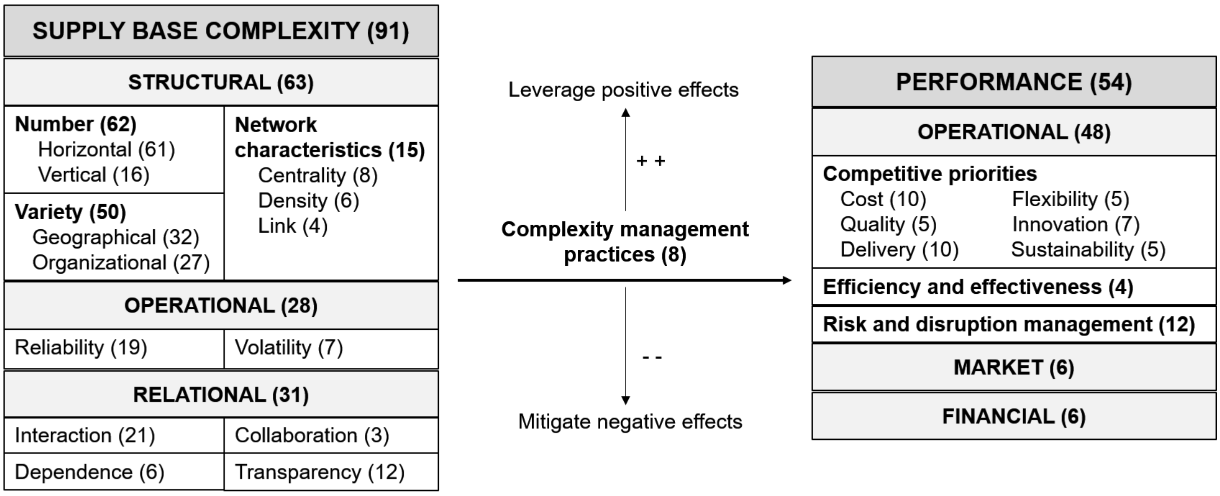
Figure 2: Research methods in SBC research



RESULTS: SUPPLY BASE COMPLEXITY DIMENSIONS

Based on our systematic literature review, we identify three main SBC dimensions: *structural*, *operational*, and *relational* (see Figure 3). The first two dimensions, *structural* and *operational* complexity, are also prevalent in complexity literature in general, whereas the latter dimension, *relational complexity*, emerges as a new complexity dimension in relation to supply chain and inter-organizational relationships perspectives. More details regarding the studies addressing each complexity dimension and performance implications are provided in the Appendix. Finally, we also evaluate the recent research stream investigating how SBC can be managed.

Figure 3. SBC dimensions and performance implications framework



Structural Supply Base Complexity

Structural supply base complexity (SSBC) consists of *number, variety,* and *network characteristics* sub-dimensions. Other terms that have been used to define structural complexity are 'static' and 'detail' complexity (e.g. Cheng et al., 2014; De Leeuw et al., 2013; Serdarasan, 2013).

One of the most discussed SSBC dimension is the *numerousness* in the system. To define numerousness, it is important to distinguish between 'horizontal' versus 'vertical' complexity. This distinction has also been used in the marketing literature, for instance in relation to product portfolios (e.g. Draganska and Jain, 2005). Horizontal complexity refers to the number of different entities in the same level whereas vertical complexity refers to the number of levels in the system (Bode and Wagner, 2015; Birkie and Trucco, 2020; Busse et al., 2017; Choi and Hong, 2002). In the upstream supply chain, horizontal complexity is defined as the number of suppliers a firm has whereas the number of supplier tiers constitutes vertical complexity. A higher number of suppliers and supplier tiers is argued to create more complexity due to increased administrative burden (Choi and Krause 2006) and additional relationships and information flows to manage (Brandon-Jones et al., 2014), ultimately creating chaos (Wilding, 1998). Numerousness appears in 62 articles accounting for approximately 68% of the total articles examined. Our results show that in majority of the studies examining numerousness, the focus is on horizontal complexity (with 61 articles) and only in 16 articles vertical complexity is discussed. Number of suppliers is often investigated with a "supply base rationalization" motive, with the implicit proposition that it is better for companies to have a smaller supply base.

The next popular SSBC dimension is *variety*, appearing in 50 articles. Variety can be defined as the difference across elements in a system (Jacobs, 2013). We see that the main variety factor investigated is how dispersed the suppliers are in terms of 'geographical distance', often defined as spatial complexity (32 studies). In 27 studies, other 'organizational' factors that can create heterogeneity in the supply base such as firm size, industry, organizational culture and technological capabilities are discussed (e.g. Ateş et al., 2015; Awaysheh and Klassen, 2010; Brandon-Jones et al., 2015; Choi and Krause, 2006; Tachizawa and Wong, 2015). Some of these studies also discuss variety/supplier heterogeneity in a broad sense and do not indicate a specific dimension.

We label the third SSBC dimension as *network characteristics*, appearing in 15 studies. Three key network characteristics are 'links/interconnectedness' (e.g. Bellamy et al., 2014), 'centrality' (e.g. Carnovale and Yeniyurt, 2015; Choi and Hong, 2002), and 'density' (e.g. Autry and Griffis, 2008; Kim and Davis, 2016). Links/interconnectedness can be defined as the degree to which supply network partners of a focal firm are connected (Bellamy et al., 2014; Lu and Shang, 2017). We find that four studies discuss links/interconnectedness in general, whereas eight studies specifically focus on centrality, the number of incident ties of a firm, (Bellamy et al., 2020; Carnovale and Yeniyurt, 2015; Tachizawa and Wong, 2015) and four studies focus on density, the number of total ties in a network relative to the number of potential ties (Tachizawa and Wong, 2015). We should note that studies discussing network characteristics do not always adopt a complexity perspective, but conceptualize these characteristics as structural dimensions. Higher number of links and density can increase SBC in terms of joint decision-making, yet result in more access to information and resources (Bellamy et al., 2014; Carnovale and Yeniyurt, 2015).

Operational Supply Base Complexity

Operational supply base complexity (OSBC), also referred to as dynamic complexity, relates to uncertainty, time, and randomness aspects of a supply base (Aitken et al., 2016; De Leeuw et al., 2013; Cheng et al., 2014; Serdarasan, 2013). The two most frequently discussed OSBC dimensions are *reliability* and *volatility*.

We find that 19 articles in our sample examine supplier delivery 'reliability'. Issues about supplier delivery reliability often stem from long lead-times and poor quality (Bozarth et al., 2009; Brandon-Jones et al., 2014; Van der Vaart et al., 2012). Suppliers, which cause shipment rejections and parts to be reworked due to low process capability, increase uncertainty not only for the buying firm, but also for the other suppliers in the network (Bozart et al., 2009; Vachon and Klassen, 2002). Long lead-times also necessitate further data and increase planning times, thereby causing operational complexity (Brandon-Jones et al., 2014; Kavilal et al., 2018; Van der Vaart et al., 2012). 'Volatility' refers to changes over time and instability in the supply base/market. Blome et al. (2014) investigate supply market dynamism suggesting that actors might change, either frequently or over time. Similarly, Johnsen et al. (2019) identify lifecycle (i.e. suppliers change over time) as a dimension of supply network complexity. A related view is presented by Ateş et al. (2015) who suggest that contract duration with suppliers (i.e. short- versus long-term contracts) can have implications for instance in terms of cost and innovation performance. Similarly, Stock et al. (2000) distinguish between one-time or ongoing interactions. A supply base consisting of mostly short-term relationships can increase SBC, as there is a constant search for new suppliers (Ateş et al., 2015; Noori and Gorgescu, 2008).

Relational Supply Base Complexity

While SSBC is related to the number of suppliers at different tier levels and the ways they are connected with the focal firm and amongst each other (Autry and Griffis, 2008), relational supply base complexity (RSBC) focuses on the nature and characteristics of these connections. We define four key dimensions of RSBC: *interaction, collaboration, transparency,* and *dependence.*

We find that 21 studies discuss interactions between actors (i.e. supplier-supplier, or supplier-buyer). It is generally argued that SBC increases if the suppliers of a focal firm interact with each other to a high extent. A high number of interactions create greater operational load borne by the focal company, especially related to managing the conflicting objectives across network members (Choi and Krause, 2006; Lu and Shang, 2017; Tachizawa and Wong, 2015). In addition to the extent of interaction, we find that three studies in our sample focus on the type of interaction (i.e. collaboration vs. competition). Whereas several supplier-supplier relationships are competitive in nature, it is not uncommon to find collaborative supplier-supplier relationships as well, for instance in large, joint new product development projects (Ateş et al., 2015; Choi and Krause, 2006). It is not extremely clear though, whether collaborative or competitive relationships create more complexity.

Six studies examine dependence/power as another supply base structure/complexity factor. Dependency can be defined as the degree to which the success of a firm in a relationship depends on the actions, resources, and capabilities of other firms (Awaysheh and Klassen, 2010; Stock et al., 2000). Finally, 12 studies highlight transparency and information sharing as another factor affecting supply base complexity. The higher the extent to which firms in a network relationship share information regarding production processes, technology, or costs, the lower the supply base complexity and the higher the supply base visibility (Kavilal et al., 2018; Stock et al., 2000). Bellamy et al. (2014) argue that even if actors do not have a direct relationship, still they could effectively access information and knowledge from other actors they are indirectly connected to.

Skilton and Robinson (2009) state that some relationships are well-defined and rigidly controlled whereas other relationships are less structured and the focal firm has little influence. Johnsen et al. (2019) also consider controllability as a dimension of supply network complexity, by focusing on whether there multiple actors controlling the network. It could be that there are many interactions between suppliers, yet if these relationships are controlled and managed in a strategic way (Aitken et al., 2016) then RSBC effects will be perceived much less. Therefore, the way interrelationships are managed should also be examined in relation to the extent of RSBC.

RESULTS: SUPPLY BASE COMPLEXITY IMPACT ON PERFORMANCE

In order to assess the implications of SBC, we examine three key dimensions of firm performance: *operational, market*, and *financial.* We find that out of 54 articles examining performance implications, the majority focus on operational performance, and we further distinguish between three sub-categories as ‘competitive priorities’, ‘efficiency and effectiveness’, and ‘risk and disruption management’. The detail about the studies addressing each dimension is shown in the Appendix. Below we elaborate on the performance implications of SBC.

Operational Performance

A large portion of the articles in our systematic literature review focuses on competitive priorities and their successful execution to define firm performance. We examine six dimensions of competitive priorities: the traditional competitive priorities of *cost, quality, delivery,* and *flexibility* (Frohlich and Dixon, 2001; Hayes and Wheelwright, 1984; Skinner, 1969), and more recently discussed competitive priorities of *innovation* and *sustainability* (Luzzini et al., 2015a, 2015b;

Martín-Peña and Díaz Garrido, 2008). We also elaborate on studies that investigate *overall operational performance*.

Competitive Priorities – Cost

We find that in ten studies the link between SBC and cost performance have been examined, and there are mixed results.

Several studies suggest a negative link between SBC and cost performance. In their conceptual study, Choi and Krause (2006) argue that SBC increases transaction costs. Interestingly, they note that to prevent this negative outcome, often companies focus on reducing the number of suppliers, but neglect other SBC dimensions such as variety and supplier-supplier interaction. Examining low cost sourcing in China, Frederiksson and Jonsson (2009) point out one of the reasons for negative impact on cost such as increased work-in-process and tied-up capital stems for longer supplier lead times. Similarly, Manuj and Sahin (2011) propose that a higher number of suppliers, an important dimension of SBC, is associated with higher transaction and total costs.

Although the main argument seems to be that high SBC results in lower cost performance, some studies illustrate that this effect is dependent on some contingencies such as the type of SBC, direct and indirect effects of SBC, and the unit of analysis. For instance, based on a large-scale survey Bozarth et al. (2009) find that delivery lead times of suppliers have a negative impact on plant cost performance, yet the number of suppliers and unreliability of suppliers are not associated with cost performance. When SBC is examined as a moderator on the relationship between supply chain integration and cost performance, Gimenez et al. (2012) find that it has no effect. By means of a multiple case study conducted in food and beverage industry, Ateş et al. (2015) highlight that the effect of SBC on cost performance is contingent on the characteristics of the purchase category. For instance, while a lower number of suppliers was associated with higher cost performance in strategic purchase categories, for non-critical purchase categories this effect was not evident.

Competitive Priorities – Quality

In contrast to cost performance, quality implications of SBC have only been briefly mentioned in the literature, and only as part of the overall supply chain complexity construct. Although Manuj and Şahin (2011) state inconsistent incoming material quality as one of the most frequently mentioned supply chain complexity outcomes, our systematic review found only a few studies examining quality performance. The focus in defining complexity in these studies seems to be on the length of supply chain and supplier lead times. Persson and Olhager (2002) argue that in simplified supply chains where lead times are shorter, there is faster feedback and therefore poor quality can be detected much more quickly. Similarly, Frederiksson and Jonsson (2009) found that product quality level is higher in shorter supply chains.

Competitive Priorities – Delivery

Similar to cost performance, delivery performance is also investigated in ten studies in our sample. Manuj and Sahin (2011) propose that a higher number of suppliers is associated with poor outbound delivery performance. Danese (2013) find that fast supply network structures with shorter supplier lead times (lower OSBC) improve buying firm schedule attainment due to shorter and more reliable production plans resulting in fewer forecast errors.

Bozarth et al. (2009) find that the number of suppliers do not have an impact on delivery performance, but delivery complexity (i.e. long lead times, unreliable suppliers) has a negative impact on schedule attainment. While it is important to distinguish between types of SBC, it is also noteworthy to distinguish between types of delivery outcomes. In their study about the impact of supply chain complexity on delivery performance, Vachon and Klassen (2002) investigate four types of delivery outcomes: throughput time, lead-time, percentage of late deliveries, and average tardiness, and find that SBC only impacts tardiness. They further argue that different forms of complexity do not have an equally negative impact on delivery performance, and even particular forms may offset others.

Competitive Priorities – Flexibility

Five studies in our sample discuss flexibility outcomes. Focusing on structural supply base characteristics, Frederiksson and Jonsson (2009) find that a high number of low-cost suppliers located abroad not only reduces cost and quality performance, but also reduces flexibility. Similarly, analyzing data from HPM survey project across several countries, Danese (2013) find that fast supply network structures with shorter supplier lead times improve buying firm flexibility through suppliers' ability to respond quickly to changes. The author states that one way to have a fast supply network structure is adopting local suppliers rather than global, suggesting that SSBC and OSCM are related. In a similar vein, Oke (2003) argues that to cope with lower delivery performance implications of more complex supply chains (caused by high number of components and suppliers) firms need volume flexibility strategies. Investigating a wide selection of sectors in Germany via a survey, Blome et al. (2014) find that SBC acts as a moderator and reduces the impact of internal knowledge transfer on supply chain flexibility. Choi and Krause (2006) state that high SBC hinders the focal company's ability to work closely with the buying firm and therefore reduce supplier responsiveness, also a dimension of flexibility. However, they also point out that different dimensions of SBC pose different levels of impact, where the negative impact of a high number of suppliers is larger than a high level of supplier-supplier interaction and differentiation between suppliers.

Competitive Priorities – Innovation

Seven studies examine the impact of SBC on innovation performance, yet report varying effects. Supporting a positive effect, Bellamy et al. (2014) show that supply network accessibility (the speed and effectiveness of information and knowledge access opportunities between a firm and its supply network) improves a firm's innovation output. In contrast, in their conceptual paper, Choi and Krause (2006) propose that there is a negative quadratic relationship between SBC and supplier innovation. Choi and Krause (2006) argue that a high number of suppliers, with different backgrounds would increase the chances of the buying firm to have more innovations; however, after some point, too much interaction between the suppliers might also cause disintegration of coherent activities and harm innovation. A potential remedy to highly interacting supply network could be to have a mix of relationships, consisting of both weak and strong ties (Autry and Griffis, 2008). Examining Italian footwear industry, Delbufalo (2015) reports supporting evidence. She finds that there is an inverted U-shaped relationship between supply base diversity (organizational variety stemming from the extent to which the competences and skills owned by the suppliers are different than the focal firm) and innovation capabilities. However, she finds a direct positive impact for the link between supply network density (interconnected networks with strong ties) and innovation.

Similarly, Sharma et al. (2020) also identify a nonlinear relationship (with diminishing growth), of other dimensions of SBC (i.e. horizontal and vertical complexity) with respect to innovation performance. In contrast, using secondary data Dong et al. (2020) find that all SBC dimensions (i.e. horizontal, vertical, spatial) attenuates the relationship between supply base R&D and financial performance, supporting more the view that SBC is detrimental.

Competitive Priorities – Sustainability

Although there has been a growing interest in sustainable supply chain management in the past decade, interestingly in our sample there is only five studies investigating the link between SBC and sustainability performance. In their conceptual paper, Tachizawa and Wong (2015) propose SBC as a moderator on the relationship between green supply chain management governance mechanisms and environmental performance, assuming a negative role for formal mechanisms and a positive role for informal mechanisms. Similarly, Jeble et al. (2018) also predict a moderating role of SBC, on the relationship between big data analytics capability and sustainability performance; however, fail to find an effect. Based on multiple case studies, Meinlschmidt et al. (2018) argue that higher supply network complexity, especially horizontal complexity, increases sustainability risk due to less visibility and higher uncertainty regarding supplier sustainability conduct.

Efficiency and Effectiveness

There are mixed results regarding the impact of supply base complexity on efficiency and effectiveness. Caridi et al. (2010) only conceptually discuss that supply chain complexity is a contingency factor affecting the impact of supply chain visibility on operational efficiency and planning effectiveness. Kao et al. (2017) examine a direct link, and argue that supply network structure (defined in terms of various network centrality measures) impact productive efficiency not only at the firm level, but also at the supply chain level. On the other hand, adopting a different conceptualization of supply base structure, Defee et al. (2010) found that only information availability (transparency) positively impacted efficiency, and there was no impact on effectiveness. A few studies found a contingency effect rather than a direct link; such as the moderating role of a fast supply network structure on the relationships between supplier integration and efficiency (Danese, 2013) and customer integration and efficiency (Danese and Romano, 2013).

Risk and Disruption Management

A total of 12 studies in our sample examine how SBC impacts risk management performance, especially focusing on supply chain disruption, resilience, and transparency. It is interesting to note that the majority of the articles that focus on supply chain disruptions was published in the past few years.

Manuj and Mentzer (2008) propose that SBC (as part of supply chain complexity) decreases the performance effects of supply chain risk management strategies due to problems regarding ownership of inventory and increased bullwhip effect. Choi and Krause (2006) extend this view by suggesting that not only high levels, but also very low levels of SBC increases vulnerability of focal firms to supply risk. Using a computational modelling approach, Basole and Bellamy (2014) find that there is a strong association between supply network structure and risk diffusion. However, one has to be cautious about interpreting this link, because supply risk is a broad concept and can increase due to a variety of reasons such as supplier delivery unreliability, limited number of suppliers, or supply disruption due to unexpected events like fire (Choi and Krause, 2006; Manuj and Dittman, 2011). Increasingly, the focus on supply risk management seems to be shifting to managing supply chain disruptions (Bode and Wagner, 2015; Brandon-Jones et al., 2015; Haberman et al., 2015).

Being one of the earliest studies, Craighead et al. (2007) state that supply chain complexity increases the severity of supply chain disruptions. In their conceptual paper, Ellis et al. (2011) adopt the conceptualization of Choi and Krause (2006), and point out that SBC is positively associated with the level of equivocality inherent in the supply disruption risk decision-making process. Based on a large-scale survey conducted across European purchasing managers, Bode and MacDonald (2016) find that when supply chain network is complex, due to ambiguity in the information being processed, there is slower response in the early stages of disruption recovery. Therefore, SBC not only increases supply disruption risk, but also impacts the effectiveness of risk management strategies. Swierczek et al. (2016) on the other hand, focus on the span of supply chain integration (i.e. basic, extended, ultimate supply chain structure), and illustrate the negative impact on the strength of the “snowball effect” in the transmission of disruptions.

Increasingly, studies also point out the difference between SBC dimensions. Using modelling, Han and Shin (2016) focus on the links between supply network members, which they argue to be a strong determinant of robustness against supply chain disruption. Bode and Wagner (2015) find that horizontal, vertical, and spatial complexity increase supply chain disruptions, and further interact and amplify each other’s effects. Similarly, Brandon-Jones et al. (2015) find that horizontal complexity and delivery complexity increase supply chain disruptions. However, their results show that spatial complexity (geographic dispersion) and supplier heterogeneity do not have a significant effect. Instead of a direct link, Brandon-Jones et al. (2014) examine the moderating role of SBC, on the relationship between supply chain visibility and supply chain resilience/robustness. They only find an effect for the number of suppliers, but not other dimensions of complexity. Habermann et al. (2015) focus on spatial complexity, and find that co-location with suppliers reduces disruption duration. They further conclude that among supply chain complexity factors, upstream factors have the largest impact, thus they caution that in order to increase supply chain resilience, more attention should be paid to SBC. The nuanced effects are also highlighted by a recent study by Wiedmar et. Al. (2021) who find that some aspects of supply network complexity intensify disruption impact, whereas others enhance disruption recovery. Chowdhury et al. (2019) argue that supply network complexity strengthens the effect of supply chain resilience on operational performance, as it can improve flexibility and reduce supply chain vulnerability. In sum, findings point out the need for differentiating between both SBC dimensions and the different phases of supply chain disruptions.

Overall Operational Performance

There were a few studies in our sample that did not explicitly distinguish between operational performance dimensions. Among them, three studies investigate SBC as an important moderator on the relationship between supply chain integration and operational performance. Gimenez et al. (2012) and Van der Vaart et al. (2012) find that for several supply chain integration practices the moderating effect is visible. On the other hand, based on the IMSS survey conducted in several countries and industries, Vanpoucke et al. (2014) argue that while the number of suppliers has a moderating effect weakening the relationship between supplier integration and operational performance, there was no effect of supply internationalization (supplier variety).

Another study by Brandon-Jones et al. (2015) do not test a direct effect of SBC, but treat it as a control variable while investigating the effect of supply chain disruptions on operational performance. Based on a large-scale, multi-industry survey conducted in UK, they find that only the number of suppliers had a negative effect, yet geographical dispersion, delivery complexity, and differentiation were not related to operational performance. Sengupta et al. (2006) examine a direct effect, but fail to find a relationship between supply network structure on operational performance both in manufacturing and service firms. They define supply network structure rather broadly by conceptualizing it as consisting of outsourcing, supplier selection, and supply base rationalization, which could potentially offset each other's effect.

An interesting remark is made by Aitken et al. (2016) who distinguish between dysfunctional and strategic complexity. They state that firms should not try to reduce complexity, but rather consider both absorption and reduction approaches for different dimensions. Based on an in-depth, single case study, they illustrate that supplier delivery complexity was successfully handled by an absorption strategy that in the end improved operational performance.

Market Performance

It is interesting to note that, compared to operational performance there is only a handful of studies focusing on market performance. Hur et al. (2004) very broadly discuss that supply chain structure in general affects customers' value and satisfaction, but do not empirically test this effect. Gerschberger et al. (2017) find that supplier-induced complexity is associated with a higher number of complaints. Finally, Bozarth et al. (2009) find that long supplier lead-times negatively affect customer satisfaction and competitive performance, but number of suppliers, unreliable supplier delivery, and global sourcing have no effect.

Financial Performance

Similar to market performance, there are very few papers examining the impact of SBC on financial performance. Sengupta et al. (2006) were not able to find a link between supply network structure and financial performance (company's relative cost and profit-related performance compared with its direct competitors). Kim and Davis (2006) did not hypothesize a direct link between SBC (number of suppliers and geographical concentration), but based on the correlations between SBC and ROA, again there was no significant effect. On the other hand, Carnovale and Yeniyurt (2015) found that network centrality increases financial performance (ROA, ROI, ROE, and ROS), but after some point these gains diminish. They also found that the more remote the firms are the poorer financial performance they have. Clearly, there is a need for more research examining financial implications of SBC.

RESULTS: MANAGING SUPPLY BASE COMPLEXITY – A RECENT RESEARCH STREAM

In their influential conceptual paper about SBC, Choi and Krause (2006) already draw attention to the issue of managing complexity by stating "Whether contemporary supply managers explicitly think in terms of supply base complexity, or not, we propose that the degree of supply base complexity affects transaction costs, supply risk, supplier responsiveness, and supplier innovation" (p. 638). Interestingly, since then the majority of the literature on SBC focused on investigating the outcomes of SBC; however, it is important to also assess whether firms take into account and manage SBC or not. Recently, there have been growing interest in understanding this issue. Being among the first studies, Aitken et al. (2016) differentiate between two practices to manage SBC: *absorbing* practices versus *reduction* practices. They suggest that not all aspects of SBC are detrimental, especially if they support the business strategy. Therefore, they argue that when there is dysfunctional complexity, reduction practices can be adopted; however, for strategic complexity it is better to implement absorbing practices. In a similar vein, Turner et al. (2018) find that instead of reducing, companies in fact mostly accommodate complexity. Adopting an ambidexterity perspective, they further discuss complexity management practices related to three main areas: *planning and control, relationship management,* and *flexibility*. Fernández Campos et al. (2019) propose a classification of complexity management practices consisting of four types: *variety reducing, confinement and decoupling, coordination and collaboration* and *decision support and knowledge generation*. These studies suggest that, if there are adequate practices in place, both structural and dynamic complexity in the supply network can be managed (Fernández Campos et al., 2019) and negative effects can be reduced (Giannoccaro et al., 2018). In a recent study, Ateş and Memiş (2021) illustrate that firms with a strategic purchasing function can mitigate the negative effects of SBC while leveraging its positive effects.

It should also be noted that too much of a control can have counter-productive effects (Giannoccaro et al., 2018). While managing SBC, managers should also take into account the complex interplay between complexity dimensions (Dittfeld et al., 2018). Another option is to adopt practices that enable partially or fully transferring complexity to suppliers (Huaccho Huatuco et al., 2020). Clearly, there is need for more research aiming to investigate which SBC management practices are effective under different contexts (Aitken et al., 2021).

DISCUSSION AND FUTURE RESEARCH AVENUES

Our systematic literature review illustrates the growing interest in SBC research. Although there is an increasing body of literature, our results show that there are several areas that require more investigation. Below, we discuss future research avenues in terms of the measurement and the structure of SBC, i.e. the content and meaning of SBC, and the role it plays in relation to its antecedents and outcomes. These two broad research questions can be split into sub-questions depending on the scope of the study (i.e., exploration, theory building, and theory testing) as shown in Table 2.

Measurement

Exploration

The literature discusses several aspects that compose the definition of SBC and we believe there is less need of exploring the meaning of SBC compared to other potential research questions. Yet, some elements of SBC might still be uncovered. For example, a growing body of literature examine the network characteristics concurring at SBC (i.e. Carnivole and Yeniyurt, 2015; Kao et al., 2017; Lu and Shang, 2017; Tachizawa and Wong, 2015). However, the network literature showcases a wide variety of measures (density, centrality and connectedness are only the peak of iceberg), that are sometimes overlapping or only slightly different. Future research might establish the fundamental network dimensions of SBC and – for the sake of parsimony – identify some measure of reference for researchers. Similarly, supplier heterogeneity is often examined in terms of geographical distance. But what about other factors that create heterogeneity such as firm size, industry, or different skill levels? Exploratory research can serve the purpose to fix the key drivers of SBC.

Table 2. Research agenda for SBC

Contextual factors

	Measurement *i.e. What is SBC?*	**Structure** *i.e. What is the role of SBC?*
Exploration	*What are the drivers of SBC? Is there any missing element of SBC?* *What is managers' cognition of SBC?*	*What are the antecedents and outcomes of SBC?* *What are the effects of SBC on market and financial performance?* *Does SBC directly affect, mediate or moderate organizational performance?*
Theory building	*What theories can help in understanding SBC?* *How are SBC dimensions related?*	*How do different SBC dimensions affect performance?* *What are the positive and negative effects of SBC?* *What are the barriers and enablers of SBC?*
Theory testing	*How can we operationalize SBC?*	*What is the net impact of SBC on performance?*

Finally, even though some studies refer to the complexity of decision-making (e.g. Manuj and Sahin, 2011; Serdarasan, 2013), there is still limited understanding about managers' cognition of SBC. Some studies emphasize that supply chain managers consider complexity as "a cancer you have to fight" (Bode and Wagner, 2015), but we know little about how managers' perceive and react to SBC.

Theory Building

Accordingly, behavioral or psychological theories can help complementing what we already know about SBC. Despite some theories have been emerging as discussed above (information entropy, complex adaptive systems, and social capital theory), other perspectives might contribute to the theory building effort on SBC. If we introduce the time dimension and consider that the supply base is evolving over time as a result of both internal and external factors to the focal firm (Lundin and Norman, 2010; Wilding, 1998), SBC acquires a dynamic nature and therefore offers further research opportunities.

Particularly relevant to this research scope would be addressing the interrelation between the different components of SBC (Dittfeld et al., 2018). As anticipated, there is some evidence in the literature that SSBC and OSBC are related, but it is not clear how and why. Moreover, even RSBC might relate to other measures. One could imagine that numerous and volatile supply bases composed of heterogeneous suppliers will be looser and based on weaker ties, whereas very dense and central supply bases will showcase stronger and more stable ties. To this regard, interpretation of possible configurations of SBC represent an interesting research stream.

Theory Testing

Finally, given that we already have possible formulations of SBC in the literature and some measurement examples of supply chain complexity (i.e. Kavilal et al., 2017), future studies might try to consolidate formulas and scales that support the operationalization of SBC.

Structure

Exploration

Despite the literature already offers some insights on possible effects of SBC, we believe there is still room for contributions regarding the outcomes of SBC. Researchers might primarily be interested in the effects on the firm's market and financial performance, which are the least explored. However, performance is not the only possible outcome variable to examine. If considered as an exogenous factor, SBC might affect – for example – the firm's purchasing strategy or the adoption of certain supply base management practices, the creation of social capital across the supply base, the level of information transparency/disclosure, the stability of relationships over time, and so on.

In the same line of reasoning, SBC does not necessarily have to be considered only as an independent variable. It would be interesting to understand which endogenous and exogenous factors determine changes in the level of SBC and if it is something managers can govern. Understanding the role of SBC on performance might also require to consider its different roles in theoretical models such as direct, moderating, mediating, and control variable, as previous studies only offer partial evidence on that.

Theory building

One of the main conclusions emerging from our synthesis is the need to examine SBC dimensions separately (Bode and Wagner, 2015; Gerschberger et al., 2017). Future research can investigate the importance/weight of each dimension in defining supply base complexity. It could be that the number of suppliers is a more decisive factor in complexity than the differentiation of suppliers. Recently, studies started to focus more on this issue, by using modelling approaches for instance (Kavilal et al., 2017). Accordingly, several studies suggest that performance implications of the dimensions also vary (Ateş et al., 2021, Brandon-Jones et al., 2014). Recent studies illustrate that SBC might be "more complex" than we think, as there are not only direct, but also curvilinear effects; both on performance, but also among the dimensions themselves (Lu and Shang, 2017). For example, Brandon-Jones et al. (2015) argue that long lead-time supply chains may also have less transparency increasing the potential for disruptions, and Dittfeld et al. (2018) illustrate that interactions between detail SSBC and OSBC complexity can reduce the overall SBC.

Aitken et al. (2016) argue that it is important to distinguish between dysfunctional versus strategic complexity. They state that if complexity is arising because of the business needs and strategy of an organization (e.g. differentiated product, requiring a more diverse supply base), then this is strategic complexity, and in that case firms that are able to absorb higher levels of complexity will gain a competitive advantage over their competitors. Manuj and Sahin (2011) bring in a different perspective, and distinguish between supply chain complexity and supply chain decision-making complexity. They explain that if there are mechanisms in place like information systems and training, although supply chain complexity is high, supply chain decision-making complexity will be lower. Therefore, one should not jump to quick conclusions regarding the negative impact of SBC; but also focus on cases where complexity might be necessary, and whether this complexity is successfully absorbed and managed. In connection to this research question, we find that the majority of the studies examining performance implications focus on cost, delivery, and disruptions. Future studies can investigate the link between SBC and less examined performance dimensions such as innovation and sustainability.

More in general, a better understanding on how SBC can affect performance requires a more in-depth view of intervening mechanisms put in place by managers that can reduce or amplify the positive and negative effects of SBC. On a related note, another interesting research avenue is to investigate how supply chain/purchasing managers cope with SBC, and how SBC affects their decision-making process in forming purchasing strategies and executing plans. Given that the ultimate results of SBC can be highly variable depending on the individual decision-makers' perception, this could be a promising research stream addressed through experiments, which is also being increasingly common in PSM field (Eckerd et al., 2016).

Theory testing

Given that the SBC literature already offers some insights and preliminary evidence on the structural paths that can exist between SBC dimensions, performance dimensions, and a few other intervening variables, there are definitely opportunities to perform some theory testing, and more will come as the field matures. For example, a relevant contribution is to provide empirical evidence about the simultaneous effect of different SBC dimensions on performance, in order to understand what might be the relative effect size. Such research effort might also support theory building in that different equivalent models might be tested, comparing mediation and moderation effects within the strategy-practice-performance paradigm.

Finally, more empirical evidence is needed regarding the effect of SBC on less common performance dimensions, particularly innovation and sustainability. On the one hand, recent studies show the popularity of open innovation paradigms and the involvement of suppliers with distant capabilities compared to the focal firm. This suggests that SBC might be tightly linked with the innovation management process, but more empirical evidence is needed to understand the consequences of different supply base structures on innovation (Sharma et al., 2020). On the other hand, sustainability research has long clarified that no company is more sustainable than its supply chain (Krause et al., 2009). Various challenges arise with increasing SBC, in term of monitoring, performance improvement, and disclosure of both the environmental and social dimensions.

Contextual Factors

Any of the research questions summarized in Table 2 and discussed above can be addressed within different contexts. Therefore, another possible future research direction is examining SBC from a contingency and fit perspective, by taking into account contextual factors such as industry, firm size, and product characteristics (Autry and Griffis, 2008). For instance, can we claim that a supply base consisting of thousand suppliers in food and beverage industry is more complex than a supply base consisting of hundred suppliers in electronics industry? Dittfeld et al. (2018) find that in their research context, food processing industry characteristics such as variability of quality and quantity of raw materials, impact the interactions between supply chain complexity dimensions.

We find that the majority of the studies in our sample focus on overall supply base as the unit of analysis, yet a few recent studies point out the need for analysing SBC also at the purchase category/product group level (Ateş et al., 2015; Lu and Shang, 2017). Especially for some SBC dimensions, a micro level unit of analysis might be more appropriate. For instance, although at the overall supply base level a firm might perceive a certain variety of suppliers, in some purchase categories suppliers might have very similar characteristics whereas in other purchase categories suppliers might be very different in terms of location, size, and so on. These differences might uncover multiple notions of SBC across the firm's purchasing portfolios.

CONCLUSION AND LIMITATIONS

In this study we conduct a systematic literature review of supply base complexity (SBC). After an overview of the main theoretical perspectives that can inform the investigation of SBC (i.e., information entropy, complex adaptive systems, and social capital), we broaden our search through a keyword-based literature review and select a final sample of 91 studies. Based on this corpus of literature, we draw a definition of SBC as composed of three main dimensions: *structural, operational,* and *relational*. Moreover, we provide an overview of the nuanced picture regarding the potential effects of SBC on *operational, market*, and *financial performance* as shown by extant studies. Furthermore, we also highlight the contingency role of *complexity management practices* that can enable mitigating the negative effects and leveraging the positive effects of SBC. Finally, we introduce some of the most relevant research questions that might offer future research opportunities.

We deem our study relevant for scholars because we provide an integrative view of SBC, which is a key premise to develop a theory of supply base management (Choi and Krause, 2006). By clarifying the different facets of SBC, we provide the necessary knowledge to better understand which kind of theories might be applied to the study of SBC, which negative and positive consequences SBC might have on performance, and which kind of intervening mechanisms might be used to hedge against or exploit SBC. Eventually, we hope to offer a starting point for the future research endeavor.

As with any other research, this study also has some limitations. First, we limit our database search to the past two decades. Second, we rely on a single database. Despite the nature of the study makes it relevant primarily for an academic audience, we believe purchasing and supply managers might find useful insights regarding what SBC is and what its effects are. Also, future follow-up studies might provide more empirical evidence regarding the most effective SBC coping strategies and management practices.

APPENDIX

		SBC DIMENSIONS													PERFORMANCE										
		STR							OPR		REL				OPR									MKT	FIN
		NUM		VAR		NET																			
		HO	VE	GE	OR	CE	DE	LI	RE	VO	IR	TR	DP	CL	CO	QU	DE	FL	IN	SU	EF	DI	OV		
Abdulsalam et al.	2015	X									X														
Aitken et al.	2016				X								X										X		
Ateş et al.	2015	X		X	X					X	X	X			X				X						
Autry and Griffis	2008						X				X								X				X		
Awaysheh and Klassen	2010			X	X							X	X												
Azadegan et al.	2013	X		X					X																X
Basole and Bellamy	2014	X				X		X														X			
Bellamy et al.	2014					X					X								X						
Birkie and Trucco	2020	X													X	X		X				X	X		
Birkie et al.	2017	X		X																		X	X		
Blome et al.	2014	X							X	X								X							
Bode and Macdonald	2016	X																				X			
Bode and Wagner	2015	X	X		X																	X			
Bozarth et al.	2009	X		X					X						X		X						X	X	
Brandon-Jones et al.	2014	X		X	X				X													X			
Brandon-Jones et al.	2015	X		X	X				X													X	X		
Busse et al.	2017	X	X	X																					
Caridi et al.	2010	X		X	X				X		X										X				
Carnovale and Yeniyurt	2015					X																			X
Chae et al.	2019	X		X							X														
Chen and Paulraj	2004	X											X												
Cheng et al.	2014	X			X					X	X														
Chettri et al.	2021				X							X		X											
Choi and Hong	2002	X	X	X	X	X																			
Choi and Krause	2006	X		X	X						X				X							X	X		
Chowdhury et al.	2019	X		X											X	X	X					X			X
Christopher	2012	X		X							X														
Craighead et al.	2007	X					X				X											X			
Danese	2013								X								X	X			X				
Danese and Romano	2013								X												X				
De Leeuw et al.	2013	X							X			X		X											
Dedrick et al	2008	X																							
Defee and Stank	2005																								X
Delbufalo	2015				X		X												X						
Dittfeld et al.	2018	X			X																				
Dong et al.	2020	X			X						X								X						
Ellis et al.	2011	X			X						X								X						
Fernández Campos et al.	2019	X			X						X		X												
Foerstl et al.	2021	X		X						X															
Frederiksson and Jonsson	2009				X										X	X		X							
Gerschberger et al.	2017			X					X															X	
Gimenez et al	2012								X						X										
Gupta et al.	2018	X		X	X					X													X	X	X
Habermann et al.	2015			X																			X		
Han and Shin	2016	X	X					X															X		
Heim et al.	2014	X							X								X								
Huang et al.	2003	X						X																X	
Hur et al.	2004	X	X	X																					
Isik	2010				X																				
Jeble et al.	2018	X			X				X											X					X
Johnsen et al.	2019			X	X					X		X													
Kao et al.	2017					X															X				
Kavilal et al.	2017	X							X						X		X								
Kavilal et al.	2018	X							X			X													
Kim and Davis	2016	X	X				X													X				X	X
Kim and Park	2008	X	X						X																
Liu et al.	2012	X																						X	

APPENDIX (continued)

		SBC DIMENSIONS													PERFORMANCE										
		STR							OPR		REL				OPR									MKT	FIN
		NUM		VAR		NET																			
		HO	VE	GE	OR	CE	DE	LI	RE	VO	IR	TR	DP	CL	CO	QU	DE	FL	IN	SU	EF	DI	OV		
Lorentz et al.	2016			X																					
Lu and Shang	2017	X	X	X				X			X														X
Lundin and Norman	2010	X	X																						
Manuj and Mentzer	2008	X	X	X								X											X		
Manuj and Sahin	2011	X													X	X	X						X		
Meinlschmidt et al.	2018	X	X	X																X					
Melville and Ramirez	2008			X							X	X													
Min and Zhou	2002	X	X																						
Oke	2003	X																X							
Persson and Olhager	2002		X												X	X	X								
Prater and Gosh	2006			X																					
Rauer and Kaufmann	2015											X	X												
Romano	2009	X	X		X									X			X								
Sengupta et al.	2006	X																					X		X
Serdarasan	2013	X			X				X		X	X	X												
Sharma et al.	2020	X	X	X															X						
Skilton	2014			X																					
Skilton and Robinson	2009	X			X						X												X		
Speier et al.	2011			X																					
Swierczek	2016																					X			
Tachizawa and Wong	2015	X			X	X	X				X									X					
Tachizawa et al.	2015	X			X	X	X				X														
Turner et al.	2018				X																				
Vachon and Klassen	2002	X		X					X	X							X								
Vachon and Klassen	2006	X																							
Van der Vaart et al.	2012								X														X		
Van der Vorst and Beulens	2002	X		X							X	X													
Vanpoucke et al.	2014	X		X																			X		
Wiedmer et al.	2021	X															X					X			
Wilhelm et al.	2016	X	X																						
Yeung et al.	2007	X									X	X													

References

Abdulsalam, Y., Gopalakrishnan, M., Maltz, A., & Schneller, E. (2015). The Emergence of Consolidated Service Centers in Health Care. *Journal of Business Logistics, 36*(4), 321–334.

Aitken, J., Bozarth, C., & Wolfgang, G. (2016). To eliminate or absorb supply chain complexity: A conceptual model and case study. *Supply Chain Management: An International Journal,* 21(6), 759–774.

Ateş, M. A., & Memiş, H. (2021). Embracing supply base complexity: The contingency role of strategic purchasing. *International Journal of Operations and Production Management,* forthcoming.

Ateş, M. A., Wynstra, F., & van Raaij, E. M. (2015). An exploratory analysis of the relationship between purchase category strategies and supply base structure. *Journal of Purchasing and Supply Management,* 21(3), 204–219.

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.

Awaysheh, A., & Klassen, R. D. (2010). The impact of supply chain structure on the use of supplier socially responsible practices. *International Journal of Operations & Production Management,* 30(12), 1246–1268.

Azadegan, A., Patel, P. C., Zangoueinezhad, A., & Linderman, K. (2013). The effect of environmental complexity and environmental dynamism on lean practices. *Journal of Operations Management,* 31(4), 193–212.

Basole, R. C., & Bellamy, M. A. (2014). Supply network structure, visibility, and risk diffusion: A computational approach. *Decision Sciences, 45*(4), 753–789.

Bellamy, M. A., Dhanorkar, S., & Subramanian, R. (2020). Administrative environmental innovations, supply network structure, and environmental disclosure. *Journal of Operations Management, 66*(7-8), 895–932.

Bellamy, M. A., Ghosh, S., & Hora, M. (2014). The influence of supply network structure on firm innovation. *Journal of Operations Management, 32*(6), 357–373.

Bhakoo, V., Singh, P. J., & Chia, A. (2015). Supply chain structures shaping portfolio of technologies: Exploring the impact of integration through the “dual arcs” framework. *International Journal of Physical Distribution & Logistics Management*, *45*(4), 376–399.

Birkie, S. E., Trucco, P., & Campos, P. E. (2017). Effectiveness of resilience capabilities in mitigating disruptions: leveraging on supply chain structural complexity. *Supply Chain Management: An International Journal, 22*(6), 506–521.

Birkie, S. E., & Trucco, P. (2020). Do not expect others do what you should! Supply chain complexity and mitigation of the ripple effect of disruptions. *International Journal of Logistics Management, 31*(1), 123-144.

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,* Part B, 307–316.

Bode, C., & Macdonald, J. R. (2016). Stages of supply chain disruption response: Direct, constraining, and mediating Factors for impact mitigation. *Decision Sciences, 48*(5), 836–874.

Bode, C., & Wagner, S. M. (2015). Structural drivers of upstream supply chain complexity and the frequency of supply chain disruptions. *Journal of Operations Management, 36*, 215–228.

Bozarth, C. C., Warsing, D. P., Flynn, B. B., & Flynn, E. J. (2009). The impact of supply chain complexity on manufacturing plant performance. *Journal of Operations Management, 27*(1), 78–93.

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.

Brandon-Jones, E., Squire, B., & Van Rossenberg, Y. G. T. (2015). The impact of supply base complexity on disruptions and performance: the moderating effects of slack and visibility. *International Journal of Production Research, 53*(22), 6903–6918.

Busse, C., Meinlschmidt, J., & Foerstl, K. (2017). Managing information processing needs in global supply chains: A prerequisite to sustainable supply chain management. *Journal of Supply Chain Management, 53*(1), 87–113.

Carey, S., Lawson, B., & Krause, D. R. (2011). Social capital configuration, legal bonds and performance in buyer–supplier relationships. *Journal of Operations Management, 29*(4), 277–288.

Caridi, M., Crippa, L., Perego, A., Sianesi, A., & Tumino, A. (2010). Do virtuality and complexity affect supply chain visibility? *International Journal of Production Economics, 127*(2), 372–383.

Carnovale, S., & Yeniyurt, S. (2015). The impact of supply network structure on the financial performance of the firm. *Supply Chain Forum: International Journal, 16*(3), 18–28.

Chen, I. J., & Paulraj, A. (2004). Towards a theory of supply chain management: the constructs and measurements. *Journal of Operations Management, 22*(2), 119–150.

Cheng, C., Chen, T., & Chen, Y. (2014). An analysis of the structural complexity of supply chain networks. *Applied Mathematical Modelling, 38*(9-10), 2328–2344.

Choi, T. Y., Dooley, K. J., & Rungtusanatham, M. (2001). Supply networks and complex adaptive systems: control versus emergence. *Journal of Operations Management, 19*(3), 351–366.

Choi, T. Y., & Hong, Y. (2002). Unveiling the structure of supply networks: case studies in Honda, Acura, and DaimlerChrysler. *Journal of Operations Management, 20*(5), 469–493.

Choi, T. Y., & Kim, Y. (2008). Structural embeddedness and supplier management: a network perspective. *Journal of Supply Chain Management, 44*(4), 5–13.

Choi, T. Y., & Krause, D. R. (2006). The supply base and its complexity: Implications for transaction costs, risks, responsiveness, and innovation. *Journal of Operations Management, 24*(5), 637–652.

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., & Handfield, R. B. (2007). The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision Sciences, 38*(1), 131–156.

Christopher, M. (2013). Managing supply chain complexity: Identifying the requisite skills*. Supply Chain Forum: An International Journal, 13*(2), 4–9.

Danese, P. (2013). Supplier integration and company performance: A configurational view*. Omega, 41*(6), 1029–1041.

Danese, P., & Romano, P. (2013). The moderating role of supply network structure on the customer integration ‐ efficiency relationship*. International Journal of Operations & Production Management, 33*(4), 372 ‐ 393.

De Leeuw, S., Grotenhuis, R., & van Goor, A. R. (2013). Assessing complexity of supply chains: evidence from wholesalers. *International Journal of Operations & Production Management, 33*(8), 960–980.

Dedrick, J., Xu, S. X., & Zhu, K. X. (2008). How does information technology shape supply-chain tructure? Evidence on the number of suppliers. *Journal of Management Information Systems, 25*(2), 41–72.

Defee, C. C., & Stank, T. P. (2005). Applying the strategy ‐ structure ‐ performance paradigm to the supply chain environment. *The International Journal of Logistics Management, 16*(1), 28 ‐ 50.

Defee, C.C., Stank, T. P., & Esper, T. (2010). Performance implications of transformational supply chain leadership and followership*. International Journal of Physical Distribution & Logistics Management, 40(*10), 763–791.

Delbufalo, E. (2015). The influence of supply network structure on firm's multiple innovation capabilities: A longitudinal study in the fashion industry. *Management Decision, 53*(10), 2457–2476.

Dittfeld, H., Scholten, K., & Van Donk, D. P. (2018). Burden or blessing in disguise: interactions in supply chain complexity*. International Journal of Operations & Production Management, 38*(2), 314–332.

Dong, Y., Skowronski, K., Song, S., Venkataraman, S., & Zou, F. (2020). Supply base innovation and firm financial performance. *Journal of Operations Management, 66*(7-8), 768-796.

Draganska, M., & Jain, D. C. (2005). Product ‐ line length as a competitive tool. *Journal of Economics & Management Strategy, 14*(1), 1 ‐ 28.

Durach, C. F., Kembro, J., & Wieland, A. (2017). A new paradigm for systematic literature reviews in supply chain management. *Journal of Supply Chain Management, 53*(4), 67–85.

Eckerd, S. (2016). Experiments in purchasing and supply management research. *Journal of Purchasing and Supply Management, 22*(4), 258–261.

Ellis, S. C., Shockley, J., & Henry, R. M. (2011). Making sense of supply disruption risk research: A conceptual framework grounded in enactment theory. *Journal of Supply Chain Management, 47(*2), 65–96.

Fernández Campos, P., Trucco, P., & Huaccho Huatuco, L. (2019). Managing structural and dynamic complexity in supply chains: insights from four case studies*. Production Planning & Control, 30(*8), 611-623.

Fredriksson, A., & Jonsson, P. (2009). Assessing consequences of low ‐ cost sourcing in China. *International Journal of Physical Distribution & Logistics Management, 39*(3), 227 – 249.

Frohlich, M. T., & Dixon, J. R. (2001). A taxonomy of manufacturing strategies revisited. *Journal of Operations Management, 19(*5), 541–558.

Frizelle, G., & Woodcock, E. (1995). Measuring complexity as an aid to developing operational strategy. *International Journal of Operations & Production Management, 15*(5), 26–39.

Gerschberger, M., Manuj, I., & Freinberger, P. (2017). Investigating supplier-induced complexity in supply chains*. International Journal of Physical Distribution & Logistics Management, 47*(8), 688–711.

Giannoccaro, I., Nair, A., & Choi, T. (2018). The impact of control and complexity on supply network performance: An empirically informed investigation using NK simulation analysis. *Decision Sciences, 49*(4), 625-659.

Gimenez, C., van der Vaart, T., & van Donk, D. P. (2012). Supply chain integration and performance: the moderating effect of supply complexity*. International Journal of Operations & Production Management, 32*(5), 583–610.

Golicic, S. L., & Smith, C. D. (2013). A meta ‐ analysis of environmentally sustainable supply chain management practices and firm performance. *Journal of Supply Chain Management, 49*(2), 78 – 95.

González-Benito, J. (2007). A theory of purchasing’s contribution to business performance. *Journal of Operations Management, 25*(4), 901–917.

Haberman, M., Blackhurst, J., & Metcalf, A. Y. (2015). Keep your friends close? Supply chain design and disruption risk. *Decision Sciences, 46*(3), 491–526.

Han, J., & Shin, K. (2016). Evaluation mechanism for structural robustness of supply chain considering disruption propagation*. International Journal of Production Research, 54*(1), 135–151.

Hartmann, E., & Herb, S. (2015). Interconnectedness of actor bonds in service triads–a social capital perspective. *Industrial Marketing Management, 44*, 154–165.

Hayes, R. H., & Wheelwright, S. C. (1984). *Restoring our competitive edge: competing through manufacturing* (Vol. 8). John Wiley & Sons New York, NY.

Heim, G. R., Peng, D. X., & Jayanthi, P. S. (2014). Longitudinal analysis of inhibitors of manufacturer delivery performance. *Decision Sciences, 45*(6), 1117–1158.

Horn, P., Scheffler, P., & Schiele, H. (2014). Internal integration as a pre-condition for external integration in global sourcing: a social capital perspective. *International Journal of Production Economics, 153,* 54–65.

Hult, G. T. M., Ketchen. D. J., Griffith, D. A., Chabowski, B. R., Hamman, M. K., Dykes, B. J., Pollitte, W. A., & Cavusgil, S. T. (2008). An assessment of the measurement of performance in international business research. *Journal of International Business Studies, 39*(6), 1064–1080.

Huang, G. Q., Lau, J. S. K., & Mak, K. L. (2003). The impacts of sharing production information on supply chain dynamics: a review of the literature. *International Journal of Production Research, 41*(7), 1483–1517.

Hur, D., Hartley, J. L., & Hahn, C. K. (2004). An exploration of supply chain structure in Korean companies. *International Journal of Logistics Research and Applications, 7*(2), 151–164.

Isik, F. (2010). An entropy-based approach for measuring complexity in supply chains. *International Journal of Production Research, 48*(12), 3681–3696.

Jacobs, M. A. (2013). Complexity: Toward an empirical measure. *Technovation, 33*(4–5), 111–118.

Jacobs, M. A., & Swink, M. (2011). Product portfolio architectural complexity and operational performance: Incorporating the roles of learning and fixed assets. *Journal of Operations Management, 29*(7–8), 677–691.

Johnsen, T. E., Mikkelsen, O. S., & Wong, C. Y. (2019). Strategies for complex supply networks: Findings from the offshore wind power industry*. Supply Chain Management: An International Journal, 24*(6), 872-886.

Kao, T. W., Simpson, N. C., Shao, B. B. M., & Lin, W. T. (2017). Relating supply network structure to productive efficiency: A multi-stage empirical investigation*. European Journal of Operational Research, 259*(2), 469–485.

Kavilal, E. G., Venkatesan, S. P., & Kumar, K. D. H. (2017). An integrated fuzzy approach for prioritizing supply chain complexity drivers of an Indian mining equipment manufacturer. *Resources Policy, 51*, 204–218.

Kim, D. Y. (2014). Understanding supplier structural embeddedness: A social network perspective. *Journal of Operations Management, 32*(5), 219–231.

Kim, S. W., & Park, S. (2008). Development of a three-echelon SC model to optimize coordination costs. *European Journal of Operational Research, 184*(3), 1044–1061.

Kim, Y. H., & Davis, G. F. (2016). Challenges for global supply chain sustainability: Evidence from conflict minerals reports. *Academy of Management Journal, 59*(6), 1896–1916.

Krause, D. R., Vachon, S., & Klassen, R. D. (2009). Special topic form on sustainable supply chain management: Introduction and reflections on the role of purchasing management. *Journal of Supply Chain Management, 45*(4), 18–25.

Lambert, D. M., & Cooper, M. C. (2000). Issues in supply chain management. *Industrial*

Leuschner, R., Rogers, D. S., & Charvet, F. F. (2013). A meta-analysis of supply chain integration and firm performance. *Journal of Supply Chain Management, 49*(2), 34–57.

Liu, G., Shah, R., & Babakus, E. (2012). When to mass customize: The impact of environmental uncertainty. *Decision Sciences, 43*(5), 851–887.

Longoni, A., & Luzzini, D. (2016). Building Social Capital into the Disrupted Green Coffee Supply Chain: Illy's Journey to Quality and Sustainability. In *Organizing Supply Chain Processes for Sustainable Innovation in the Agri-Food Industry* (pp. 83–108). Emerald Group Publishing Limited.

Lorentz, H., Töyli, J., Solakivi, T., & Ojala, L. (2016). The effect of a geographically dispersed supply base on downside risk: Developing and testing the N-shaped theory. *International Business Review, 25*(4), 872–882.

Lu, G. & Shang, G. (2017). Impact of supply base structural complexity on financial performance: Roles of visible and not-so-visible characteristics. *Journal of Operations Management, 53–56,* 23–44.

Lundin, J. F. & Norrman, A. (2010). The misalignment cycle: is the management of your supply chain aligned? *International Journal of Physical Distribution & Logistics Management, 40*(4), 277–297.

Luzzini, D., Amann, M., Caniato, F., Essig, M., & Ronchi, S. (2015a). The path of innovation: purchasing and supplier involvement into new product development. *Industrial Marketing Management, 47,* 109–120.

Luzzini, D., Brandon-Jones, E., Brandon-Jones, A., & Spina, G. (2015b). From sustainability commitment to performance: The role of intra-and inter-firm collaborative capabilities in the upstream supply chain. *International Journal of Production Economics, 165,* 51–63.

Manuj, I., & Menzter, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution & Logistics Management, 38*(3), 192–223.

Manuj, I., & Sahin, F. (2011). A model of supply chain and supply chain decision‐making complexity. *International Journal of Physical Distribution & Logistics Management, 41*(5), 511‐549.

Martín-Peña, M. L., & Díaz-Garrido, E. (2008). A taxonomy of manufacturing strategies in Spanish companies. *International Journal of Operations and Production Management, 28*(5), 455–477.

McIvor, R., Humphreys, P., & McAleer, E. (1997). The evolution of the purchasing function*. Strategic Change, 6*(3), 165–179.

Meinlschmidt, J., Schleper, M. C., & Foerstl, K. (2018). Tackling the sustainability iceberg. *International Journal of Operations & Production Management, 38*(10), 1888-1914.

Melville, N., & Ramirez, R. (2008). Information technology innovation diffusion: an information requirements paradigm. *Information Systems Journal, 18*(3), 247–273.

Min, H. & Zhou, G. (2002). Supply chain modeling: past, present and future*. Computers & Industrial Engineering, 43*(1-2), 231-249.

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2015). *Purchasing and supply chain management.* Cengage Learning.

Nahapiet, J., & Ghoshal, S. (2000). Social capital, intellectual capital, and the organizational advantage. In *Knowledge and social capital* (pp. 119–157). Elsevier.

Noori, H., & Georgescu, D. (2008). Making supply chain design the rational differentiating characteristic of the OEMs. International Journal of Production Research, 46(10), 2765–2783.

Oke, A. (2003). Drivers of volume flexibility requirements in manufacturing plants. *International Journal of Operations & Production Management, 23*(12), 1497–1513.

Persson, F., & Olhager, J. (2002). Performance simulation of supply chain designs. *International Journal of Production Economics, 77*(3), 231–245.

Prater, E., & Ghosh, S. (2006). A comparative model of firm size and the global operational dynamics of U.S. firms in Europe. *Journal of Operations Management, 24*(5), 511–529.

Rauer, J., & Kaufmann, L. (2015). Mitigating external barriers to implementing green supply chain management: A grounded theory investigation of green-tech companies' rare earth metals supply chains. *Journal of Supply Chain Management, 51*(2), 65–88.

Roden, S., & Lawson, B. (2014). Developing social capital in buyer–supplier relationships: The contingent effect of relationship-specific adaptations. *International Journal of Production Economics, 151,* 89–99.

Romano, P. (2009). How can fluid dynamics help supply chain management? *International Journal of Production Economics, 118*(2), 463–472.

Sengupta, K., Heiser, D. R., & Cook, L. S. (2006). Manufacturing and service supply chain performance: A comparative analysis. *Journal of Supply Chain Management, 42*(4), 4–15.

Serdarasan, S. (2013). A review of supply chain complexity drivers. *Computers & Industrial Engineering, 66*(3), 533–540.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Sharma, A., Pathak, S., Borah, S. B., & Adhikary, A. (2020). Is it too complex? The curious case of supply network complexity and focal firm innovation. *Journal of Operations Management, 66*(7-8), 839-865.

Simon, H.A. (1962). The architecture of complexity*. Proc. Am. Philos. Soc., 106*(6), 467–482.

Skilton, P. F. (2014). Value creation, value capture, and supply chain structure: Understanding resource–based advantage in a project–based industry. *Journal of Supply Chain Management, 50*(3), 74–93.

Skilton, P. F., & Robinson, J. L. (2009). Traceability and normal accident theory: How does supply network complexity influence the traceability of adverse events? *Journal of Supply Chain Management, 45(*3), 40–53.

Skinner, W. (1969). Manufacturing-missing link in corporate strategy. *Harvard Business Review, 47*(3), 136–144.

Speier, G., Whipple, J. M., Closs, D. J., Voss, M. D. (2015). Global supply chain design considerations: Mitigating product safety and security risks. *Journal of Operations Management, 29* (7-8), 721–736.

Stock, G. N., Greis, N. P., & Kasarda, J. D. (2000). Enterprise logistics and supply chain structure: the role of fit. *Journal of Operations Management, 18*(5), 531–547.

Swierczek. (2016). The “snowball effect” in the transmission of disruptions in supply chains: The role of intensity and span of integration. *The International Journal of Logistics Management, 27*(3), 1002–1038.

Tachizawa, E. M., & Wong, C. Y. (2015). The Performance of Green Supply Chain Management Governance Mechanisms: A Supply Network and Complexity Perspective. *Journal of Supply Chain Management, 51*(3), 18–32.

Tachizawa, E. M., Alvarez-Gil, M. J., & Montes-Sancho, M. J. (2015). How “smart cities” will change supply chain management. *Supply Chain Management: An International Journal, 20*(3), 237–248.

Tchokogué, A., Nollet, J., & Gobeil, K. (2011). Supply structure: The pendulum in action. *Journal of Purchasing and Supply Management, 17(*3), 148–157.

Thompson, J. D. (1967). Organizations in action: Social science bases of administrative theory. Transaction publishers.

Turner, N., Aitken, J., & Bozarth, C. (2018). A framework for understanding managerial responses to supply chain complexity*. International Journal of Operations & Production Management, 38*(6), 1433-1466

Vachon, S., & Klassen, R. D. (2002). An exploratory investigation of the effects of supply chain complexity on delivery performance. *IEEE Transactions on Engineering Management, 49*(3), 218–230.

Vachon, S., & Klassen, R. D. (2006). Extending green practices across the supply chain: The impact of upstream and downstream integration. *International Journal of Operations & Production Management, 26*(7), 795–821.

Van der Vaart, T., van Donk, D. P., Gimenez, C., & Sierra, V. (2012). Modelling the integration ‐ performance relationship: Collaborative practices, enablers and contextual factors. *International Journal of Operations & Production Management, 32*(9), 1043 ‐ 1074.

Van der Vorst, J. G. A. J., & Beulens, A. J. M. (2002). Identifying sources of uncertainty to generate supply chain redesign strategies. *International Journal of Physical Distribution & Logistics Management, 32*(6), 409–430.

Vanpoucke, E., Vereecke, A., & Wetzels, M. (2014). Developing supplier integration capabilities for sustainable competitive advantage: A dynamic capabilities approach. *Journal of Operations Management, 32*(7–8), 446–461.

Wiedmer, R., Rogers, Z. S., Polyviou, M., Mena, C., & Chae, S. (2021). The dark and bright sides of complexity: A dual perspective on supply network resilience. *Journal of Business Logistics*, forthcoming.

Wilding, R. (1998). The supply chain complexity triangle: Uncertainty generation in the supply chain. *International Journal of Physical Distribution & Logistics Management, 28(8),* 599–616.

Wilhelm, M., Blome, C., Wieck, E., & Xiao, C. Y. (2016). Implementing sustainability in multi-tier supply chains: Strategies and contingencies in managing sub-suppliers. *International Journal of Production Economics, 182,* 196–212.

Yeung, J. H. Y., Selen, W., Deming, Z., & Min, Z. (2007). Postponement strategy from a supply chain perspective: cases from China. *International Journal of Physical Distribution & Logistics Management, 37(*4), 331–356.

DECISION SCIENCES INSTITUTE

No Panic in Pandemic: The Impact of Individual Choice on Public Health Policy and Vaccine Priority

Miao Bai
Department of Operations and Information Management, University of Connecticut
miao.bai@uconn.edu

Ying Cui
Department of Industrial and Systems Engineering, University of Minnesota
yingcui@umn.edu

Guangwen Kong
Fox Department of Marketing & Supply Chain Management, Temple University
guangwen.kong@temple.edu

Anthony Zhenhuan Zhang
Department of Industrial and Systems Engineering, University of Minnesota
zhan4490@umn.edu

ABSTRACT

We study strategic planning of public health interventions (social distancing and lockdown) in pandemics by considering individual response. We derive structural properties of optimal interventions and conduct numerical studies based on Minnesota COVID-19 data. We find that the individual equilibrium activity level is higher than the socially optimal activity level due to an individual's ignorance of the negative externality imposed on others. As a result, lockdowns and social distancing are more effective when the disease prevalence is not at its peak level. Moreover, vaccination priority strategies need to consider the trade-off between mortality rates and negative externality of different groups.

<u>KEYWORDS</u>: COVID-19, Dynamic compartmental model, Public health policy analysis, and Game theory

DECISION SCIENCES INSTITUTE

Blockchain Conceptualization through Systemic Lenses

Kiran Patil
University of North Texas
Kiran.Patil@unt.edu

ABSTRACT

Blockchain is expected to alter how businesses operate since organizations have recognized its value. Existing research has examined blockchain's potential, acceptance, consequences, and use cases. However, previous research lacks unanimity on adequately conceptualizing blockchain. This study closes this gap by utilizing the systems thinking approach and the systemigram, a systemic modeling technique based on the Boardman Soft Systems Methodology. Using systemigram, this research conceptualizes blockchain as a thorough structure of complicated interlinkages that end in a single systemic purpose. This study boosts the formulation of research agenda to assist organizations in scrutinizing the enthusiasm around blockchain.

Keywords: Blockchain, Systemigram

INTRODUCTION

Blockchain is supposed to revolutionize the way organizations conduct business (Tapscott & Tapscott, 2017). In recent years, organizations have begun to realize the importance of blockchain, an open standard technology, as an alternative to closed standard systems (Risius & Spohrer, 2017). Blockchain technology is a peer-to-peer distributed asset database that can be shared across a network of multiple sites, geographies, or institutions. It can validate, record, and distribute transactions in immutable and encrypted ledgers. Organizational motivations to employ Blockchain technology include extended visibility and traceability, supply chain digitalization and disintermediation, improved data security, and smart contracts. Blockchain is better suited for dynamic supply chains and requires intense collaboration between organizations that span the globe (Wang, Singgih, Wang, & Rit, 2019). It is supposed to disrupt current industries extensively, leading to new types of companies (Felin & Wilson, 2018).

Unlike the internet, which took two decades to develop and another decade to become commercially viable, the blockchain network is growing more briskly as an economic platform. What is it that blockchain will offer to businesses to make it stand out? Blockchain technology proposes providing internet of value - a safe platform, database, or ledger where sellers and buyers can store and trade value without involving conventional mediators. Like many of the technologies, blockchain has several advantages. Primarily, this technology is distributed, which means there is no central database that can be hacked as it runs on computers provided by volunteers across the world. The second benefit, it is public since it exists in the network; anyone can access it at any time. Furthermore, the third advantage is security due to heavy-duty encryption (Tapscott & Tapscott, 2017).

Extant research has investigated blockchain from standpoints of its capabilities, adoption, influence on supply chain resilience, and use cases in sectors such as finance (Risius & Spohrer, 2017). Researchers have also studied blockchain as a mechanism to share information, applying theoretical lenses, such as transaction cost economics, contingency theory, resource-based view, resource dependency theory, and relational governance theories (Kembro, Selviaridis, & Näslund, 2014). However, past research lacks consensus on describing blockchain in its entirety. This study attempts to address this gap using systems thinking approach and systemigram, a systemic diagramming method based on Boardman Soft Systems Methodology (Sauser & Boardman, 2015).

Applying systems thinking, this study balances multiple perspectives, provides an ability to visualize the entire phenomenon, explains the mechanism behind different entities coming together in a single system, and identifies the system's emergent properties, which comprises organizations linked in a blockchain. As per systems literature, a system is a set of interconnected things to produce their pattern of behavior over time. The components of the system are interdependent and interact to form a complex and unified whole. The mutual interaction of components is crucial for the system's existence (Gandhi, Gorod, & Sauser, 2012).

A systemigram exhibits parts, wholes, emergences, boundaries, flows, inputs, outputs, transformations, processes, and networks (Sauser, Mansouri, & Omer, 2011). All these elements are fundamental to systems thinking. Systemigram is an effective methodology since it unearths the motivations, views, interactions, and qualitative dimensions of a central phenomenon (Prins, Farr, McDonald, Fitzgerald, & Sanchez, 2015; Sauser et al., 2011). The underlying principles of accuracy, relevancy, comparability, clarity, and systematic design further justify the choice of using systemigram to define blockchain (Sauser & Boardman, 2015). Using systemigram, this study characterizes blockchain as a detailed structure of complex interrelationships that culminates in a single objective of the system through exploring diverse alternative solutions.

Blockchain Conceptualization through Systemic Lenses

Initial findings suggest that blockchain creates a single shared system of records, adhering to cryptography standards (Wu et al., 2017). It results in secure inter-organizational transactions and trust, thereby creating value for organizations (Wang, Han, & Beynon-Davies, 2019). Apart from presenting a holistic view of blockchain, based on an in-depth literature review, this study contributes to setting a research agenda and brings forth research questions that aid organizations in evaluating the hype surrounding blockchain.

BACKGROUND

Blockchain

There are several strategic benefits of blockchain to corporations. First, blockchain can enable new businesses and operating models for both startups and incumbents. Furthermore, it can disrupt existing industries. Companies need to link blockchain technology with their strategy and capabilities systematically to reap the benefits and to create value using this technology. Blockchain can provide a basis for powerful applications that can simplify core operations, such as more transparency in intellectual property ownership, seamless and automated payments, improving efficiency and transparency in the supply chains (Felin & Wilson, 2018).

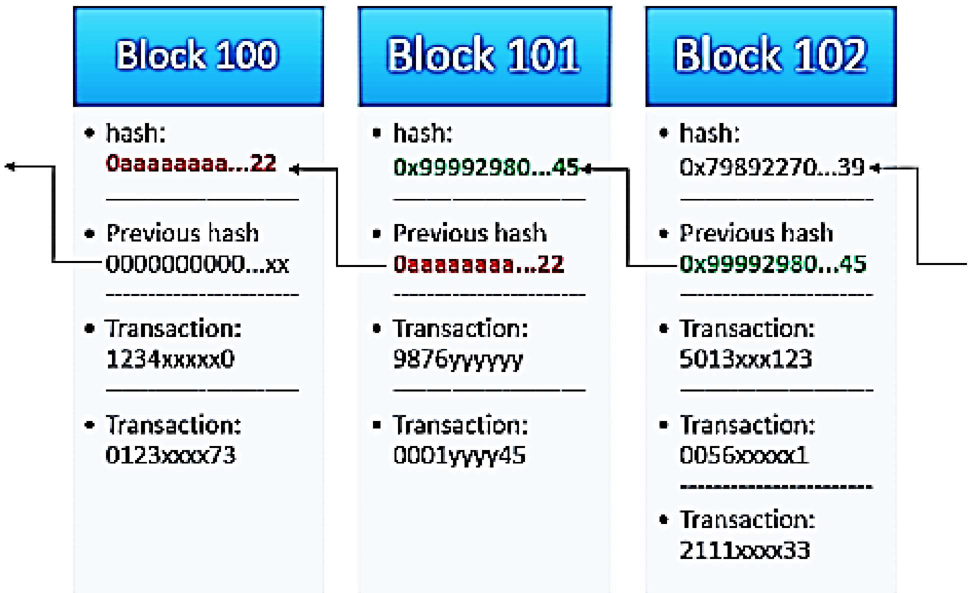


Figure 1: Blockchain mode of operations.

(Source: adapted from (Mandart, 2017) The Crypto Guide for Beginners – What is blockchain? Retrieved May 9, 2019, from CONDA AG, Donau-City-Straße 6, 1220 Vienna / Austria website: https://ico.conda.online/the-crypto-guide-for-beginners-—-what-is-blockchain)

Blockchain transactions function in a distributed way in a peer-to-peer network (Y. Chen, 2018). This function means it is not obligatory to have a central entity to validate transactions, yet the transactions are validated and stored by a distributed consensus. Fig.1 interprets the blockchain fundamentals. Fig.1 shows that each block is connected to the preceding one (forming chain). With this integration, any information history can be recovered quickly just by exploring the previous blocks. Each block has its hash with a unique ID, and it also fetches the hash of the previous block to create a more secure transaction.

Furthermore, a set of recorded and validated transactions by other computers of the same network are stored in the blocks. During this process, the transactions get a unique sequence and timestamp. The transaction cannot be altered after validation. Moreover, transparency of operations is boosted as transactions are shared across the network and any other helpful information, allowing all involved parties to get the complete information in due time. This feature simultaneously creates trust. Blockchain integrated with the supply chain field has vast potential to renovate the relationships between network members, increase efficiency and restructure transaction costs (Queiroz & Fosso Wamba, 2019).

The blockchain is a peer-to-peer transaction platform, which excludes intermediaries. Different entities involved in the transaction work as nodes, and the process is being validated through encryption. The history of these transactions is stored in the form, which covers all participating entities. Since blockchain results from the integration of several other technologies such as software development, cryptographic technology, database technology, etc., it can be called meta-technology. The traditional double-entry accounting system that came into existence in the fifteenth century has the significant disadvantage of trust-related issues. Blockchain technology eliminates this disadvantage, thereby reduces friction in the transaction process. Due to its numerous powerful benefits, blockchain is sometimes termed as next digital revolution. It is often widely evaluated with the progress of the internet in the early 90s. Considering the complex nature of modern era supply chains, blockchain is expected to speed up the processes and make them more reliable. It is contended that blockchain will bring a significant shift from 'Internet of Information' to 'Internet of Values,' making transactions quicker and inexpensive (Kamble, Gunasekaran, & Arha, 2018).

Better tracking and more transparency have a massive impact on any supply chain, and blockchain can be instrumental on this front. Kshetri (2018) studied the role of blockchain in meeting critical SCM objectives such as cost, quality, speed, dependability, risk reduction, sustainability, and flexibility using innovation diffusion theory. He found that industries with many supply layers, for example, the oil trading industry, would surface as the prime candidate in blockchain adoption. Using the case of farmers, the study predicted that the adoption of blockchain by one entity would put related pressure on other entities of the supply chain (Kshetri, 2018). Nakasumi (2017) recommends implementing a blockchain-based solution to address traditional supply chain concerns such as double marginalization and information asymmetry (Nakasumi, 2017). Korpela, Hallikas, and Dahlberg (2017) discuss how digital supply chain transformation be achieved with blockchain integration. The study proposes to bridge the gap between business readiness and current functionalities using blockchain- A cost-effective and flexible integration model based on the cloud. This study mentions ledger, smart contracts, and timestamps as the most useful functionalities of the blockchain (Korpela, Hallikas, & Dahlberg, 2017). Analyzing the causes of lack of trust in supply chain quality management, Chen et al. (2017) propose a theoretical framework for the SCQM system based on the blockchain. The system architecture includes four layers- the IoT layer, data layer, contract layer, and business layer. This model is anticipated to serve as a digital identity and help in contract automation, real-time quality monitoring, and logistics planning (Chen, R. Shi, Ren, Yan, & Zhang, 2017). Omran et al. (2017) propose a conceptual framework for a blockchain-driven supply chain finance solution. The blockchain is proposed as the most promising solution to achieve key value drivers of supply chain efficiency, transparency, and autonomy. They introduce a blockchain technology framework based on reverse factoring and dynamic discounting to improve decision-making (Omran, Henke, Heines, & Hofmann, 2017). Tian (2017) proposes a traceability system in the food supply chain for real-time tracing based on HACCP (Hazard Analysis and Critical Control Points), blockchain, and IoT. The proposed process assists in tracking and identifying fake products. It is claimed that food chains can be transformed using blockchains (Tian, 2017). Wu et al. (2017) suggest an online shipment tracking framework based on a privately distributed ledger and a single blockchain public ledger (Wu et al., 2017). Blockchain's unique solution based on combining private and public ledger address the traditional tracking system challenge of independent validation of shipment tracking information. Toyoda et al. (2017) propose a product ownership management system of RFID-attached products for anti-counterfeits based on blockchain (Toyoda, Takis Mathiopoulos, Sasase, & Ohtsuki, 2017). Abeyratne and Monafared (2016) discuss the several benefits of blockchain in different application areas of the manufacturing supply chain using cardboard boxes. The study claims that IoT and blockchain will significantly influence next-generation manufacturing (Abeyratne & Monfared, 2016) (Kamble et al., 2018).

Blockchain can influence to achieve key supply chains management objectives such as cost, quality, speed, dependability, risk reduction, sustainability, and flexibility. Blockchain has important cost-saving suggestions. Blockchain can generate code even for small transactions with minimal cost. Blockchain can help identify the source of defective products, e.g., contaminated food, and engage in tactical removals of affected products instead of recalling the entire product line. It also assists in allocating the right amount of resources to carry out activities, eliminate paper records, and provide the right indicative data to supply chain partners to describe the quality and reduce low quality and counterfeit products. Blockchain can help digitize physical processes, and a reduction in manual interactions and communications can increase the speed of the supply chain. Blockchain applications such as digitally signed documents' secure storage and transmission can make supply chain partners more accountable and responsible for their actions, creating a high level of dependability. Blockchain's traceable records can address challenges associated with self-reported data that supply chain partners provide. With the security benefits of blockchain, various current concerns can be addressed. For example, parties in the network can only access transactions, even specific touchpoints with mutual acceptance. Downloaded software file integrity can be ensured since it cannot be breached. With blockchain, it is possible to make sustainability indicators more quantifiable and more meaningful. Blockchain solutions have a powerful effect even when only a few participants use them. The power of this solution increases with the network effect. Blockchain will have a higher impact on logistics and supply chains with complementary technological advancements, such as the Internet of Things. Thus, blockchain will be instrumental in reducing supply chain costs, increasing speed, reducing risks, improving sustainability measures, improving flexibility, and reducing dependability (Kshetri, 2018).

Systems Thinking

The concept of a system is foundational to this paper. It is defined in multiple ways. First, a system is a set of things, such as people, cells, functions, that are interconnected to produce their pattern of behavior over time. Second, it is conceived as a group of interacting, interrelated, and interdependent components that form a complex and unified whole. Third, it is also defined as an entity that maintains its existence through the mutual interaction of its parts (Meadows, 2008).

A system is broadly classified into four types – mechanistic, animate, social, and ecological systems, to increase the scope encompassed. Mechanistic systems are at the granular level and are a component of all three other types of systems. Mechanistic, animate, social, and ecological systems comprise machines, people, society, and environment or worldview, respectively. Animate and social systems are purpose-driven since they can achieve the same outcome in multiple ways, in the same environment, and different outcomes in variable environments (Sauser & Boardman, 2015).

Mechanistic systems and their components have no purposes of their own, but their essential parts enable the functioning of the whole system. All mechanisms, such as plants or clocks, are mechanical systems. They operate with a regularity dictated by their internal structure and the causal laws of nature. For example, an automobile is a mechanical system that has no purpose but serves its driver and passengers' purposes. In addition, an automobile's fuel pump (a mechanical system) has the function of supplying its fuel injector or carburetor with fuel, without which the automobile will not be able to carry out its defining function. Furthermore, these systems are either open or closed, depending on the external effects on their behavior. A universe is conceptualized as a closed system, while a planet is observed as an open mechanistic system (Sauser & Boardman, 2015).

Animate systems are conceptualized as purposeful systems whose parts have no purposes of their own. The principal purpose of such systems is survival. For example, a person's lungs have no purpose but are vital to enable a person to extract oxygen from the environment to survive. These systems are necessarily open since they must interact with their environments in order to survive. Understanding these interactions is essential for understanding their properties and behavior. Also, these systems are said to contain 'life' (Sauser & Boardman, 2015).

Social Systems have their purposes, consist of animate and mechanistic parts, and are a part of one or larger encompassing system, which may have its purposes, may contain other social systems. For example, the organization is a social system by itself. When organizations are interconnected through collaborations at a network level, social systems are nested (Sauser & Boardman, 2015).

Ecological Systems contain mechanistic, animate, and social systems as parts and, therefore, some of their parts have their purposes. However, these systems have no purpose as a whole. For example, nature or the environment is understood as an ecological system. These systems serve the purposes of their animate and social parts and provide necessary inputs to these open deterministic systems. They also provide a receptacle for their waste and valuable products (Sauser & Boardman, 2015). Inanimate and social systems, choices can affect their effectiveness, either positively or negatively, thereby leading to problematic situations. Thus, problems result from the system's choice to significantly affect the system (Sauser & Boardman, 2015).

Systems thinking entails adopting the perspective of 'system' as a concept to the phenomenon of interest (Gandhi et al., 2012). A researcher is like an external observer to the system and intends to conceptualize its behavior from multiple viewpoints, starting with classifying the system as one of the four system types conceptualized above (Sauser & Boardman, 2015). The prime objective is to capture the parts of the system and how these parts fit in the whole. At the same time, the whole system's behavior is checked whether it equals the sum of behaviors of its constituent parts. The interaction of the parts and their contribution to the system's behavior under observation forms the underlying key outcome of the systems thinking process (Sauser & Boardman, 2015). Multiple tools can be used to apply systems thinking. Soft systems methodology focuses on describing a foundational framework for the investigation of the systems of interest (Sauser, Boardman, & Gorod, 2008). System information is organized using a conception. Systemigram is a popular tool to structure the system information such as that it is possible to conceptualize the entire system, with its components, interactions, connections, and purpose, in a single observation window. System behavior drivers and the underlying mechanism is understood using causal loops (Gonul Kochan, Nowicki, Sauser, & Randall, 2018). Systems dynamics modeling is used to quantify the system behavior and test the model using simulations (Prins et al., 2015). Role-playing is a popular and simple way to simulate the system comprising of human behaviors (Prins et al., 2015). Agent-based modeling quantifies human system behavior

(John, John, Parker, Sauser, & Wade, 2018). This study adopts Boardman's soft system methodology, using systemigram to conceptualize the intricacies of the blockchain (Sauser & Boardman, 2015).

METHODOLOGY

Boardman Soft Systems Methodology

This study adopts Boardman Soft Systems Methodology to structure and conceptualize the phenomenon of blockchain as a system. In contrast to rigid systems, which are bounded by physical existence that can be designed to accomplish a given objective, soft systems are assumed to possess flexible boundaries beyond physical existence perceptions. They are not understood as ontological entities, as is familiar with rigid systems (John et al., 2018). Soft systems are understood to follow the notion of systems as epistemological since it contains mental constructs used for human understanding. Considering an organization as a soft system, it can be conceptualized in multiple ways depending on the underlying prime objectives, ranging from profit generation, the transformation of materials, employment, to resource consumption leading to environmental hazards (Sauser & Boardman, 2015).

The wide variety of ways to conceptualize an organization as a soft system points to the depth and diversity in the application of Boardman Soft System Methodology. A researcher needs to be conscious of particular perspectives and values beneath the research agenda for practical application. These further determine the acceptability of things identified as a part of the system. For example, when an organization is viewed as a system to transform materials to suit human consumption, it will likely include material suppliers and product buyers as the vital parts of the system. On the other hand, perceiving an organization as a medium to employ people, an organization is understood to contain infrastructure, such as transportation and information, that allows the community members to access and contribute to the organizational objective. Thus, depending on the objective concerns and perceptions, a soft system has diverse boundaries (Sauser & Boardman, 2015).

The procedure to apply Boardman Soft System Methodology starts with identified the problem situation, in which the system shows significant behavioral changes under external influences. The problem is conceptualized in an unstructured way, capturing the variations in the system behavior. Then, the unstructured problem situation is expressed through linguistics. As a result, structures emerge from the information, which is evident in understanding the human intellect. The structures are then drawn using systemigram, following the prescribed guidelines. A researcher formalizes the system concept and thinking as applied to the phenomenon of interest. The findings depicted in the systemigram are opened for discussion, which leads to dramatization through tools, such as role-play, and a dialogue to understand the functioning of the system. In the process, systemigram findings are rectified as stakeholders in the discussion eliminate redundant system components and connections. Finally, actions to improve the problem situation are identified through consensus, which forms the prime objective of the system (Sauser & Boardman, 2015).

Systemigram

A systemigram is a powerful tool to conceptualize the system in a storytelling manner and is helpful to provide a common foundation to foster a dialogue during discussions of the system description. It does not eliminate complexity from the system but makes the system, as a whole, understandable to the human mind. Salient features of the system are covered in the systemigram, as lengthy documentation is distilled down to concentrated text or narrative. The problem situation is conceptualized in the narrative, followed by constructing a visual systemigram, decomposing text and narrative into individual, related threads, showing the flow of information, resources, and actions (Sauser et al., 2011).

This study adopts systemigram because of its suitable attributes to conceptualize the highly debated blockchain system. A systemigram enables underlying reasoning about the problem, understanding the problem's complexity, and explaining the problem situation in simple language. In addition, it enables accomplishing the aspects of reason, complexity, and explanation, which are fundamental objectives of modeling, from a multi-stakeholder perspective (Sauser et al., 2011).

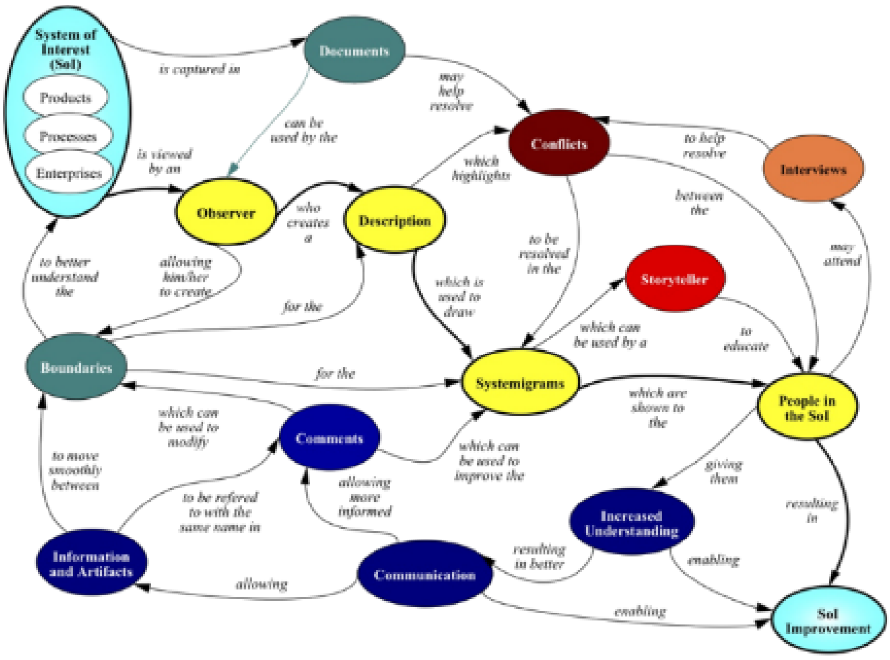


Figure 2: Systemigram of Systemigram

(Source: adapted from Sauser, B., & Boardman, J. (2015). Modeling and Simulation Support for System of Systems Engineering Applications (1st ed.; L. Rainey & A. Tolk, eds. https://doi.org/10.1002/9781118501757.ch11)

Figure 2 shows the systemigram of a systemigram to conceptualize the process employed during systemigram construction. It contains nodes and links as components and connections of the system, respectively. The starting node is the system of interest, comprising of products, processes, and enterprises. The ending node is the objective of the system, which is to improve the system. An observer, who creates a description of the system, views the system of interest. This description is used to draw the systemigram. The people involved in the system are shown the visual depiction of the understanding of the system of interest in the form of a systemigram, leading to discussions and dialogues. As ideas, thoughts, opinions, reasoning, and assumptions bounce off in a healthy discussion, the people in the system of interest identifying means to improve the system. As the systemigram is captured in documents, which may help resolve conflicts during the discussion on rectifying the systemigram, it is also used as a storytelling tool to the people in the system of interest, increasing their understanding. This systemigram further improves communication, allowing for more informed comments that modify the boundaries to understand the interest system better. People in the system of interest may attend interviews to resolve conflicts amongst them, as people in the system of interest gain better understanding, which results in improved communication, allowing information and artifacts to surface and be organized in a structure with flexible boundaries. The information and artifacts are referred to with the same terminology in the comments, which regulate the boundaries of the understanding of the system (Sauser & Boardman, 2015).

The construction of the systemigram is regulated through uniform rules to design it. The nodes must contain noun phrases, such as people, organizations, groups, artifacts, and conditions. The links should contain verbs or verb phrases that demonstrate the logical connection between the nodes. Thus, parts will tend to be nouns, or noun phrases, verbs, or verb phrases will express the relationships between the nodes, and the whole structure of systemigram is an arrangement of the parts and their relationships. The nodes cannot be repeated in a single systemigram as redundant nodes harm the essence of the relationship. Crossing over links is not permitted to keep the systemigram clear to observe critical heuristics in system design. Identifying the nodes, relationships connecting those nodes, and arranging them in a structure, which explains the system logically and coherently, is repeated until a successful outcome of Boardman Soft System Methodology is achieved. Success, here, is defined as the satisfaction of the people concerned with the system or stakeholders in improving the problem situation and gaining insights to deploy solutions (Sauser & Boardman, 2015).

Systemigram adheres to the principles that are foundational to modeling processes and outcomes. These principles include accuracy, relevance, feasibility, clarity, comparability, systematic design, and conciseness. The accuracy in the systemigram is validated through the mainstay, which supports the prime objective of the system and reads from top left to bottom right. The strict guidelines of nodes and links containing nouns and verbs, respectively, and abstinence to repeat nodes and cross-links ensure accuracy. A systemigram is always focused on the system of interests of the stakeholders, thereby remaining relevant. Since a systemigram compares the reality to the understanding of the system, it follows the principle of feasibility. Each version of the understanding of the system is repeatedly challenged to identify the rightful owner of the system and its nodes. Every node, except for the beginning and end nodes, is forced to be identified with inputs and outputs such that any node in the systemigram can be logically followed to the end node. Thus, a systemigram, based on Boardman Soft System Methodology, results from systematic design (Sauser & Boardman, 2015; Sauser et al., 2011).

RESULTS

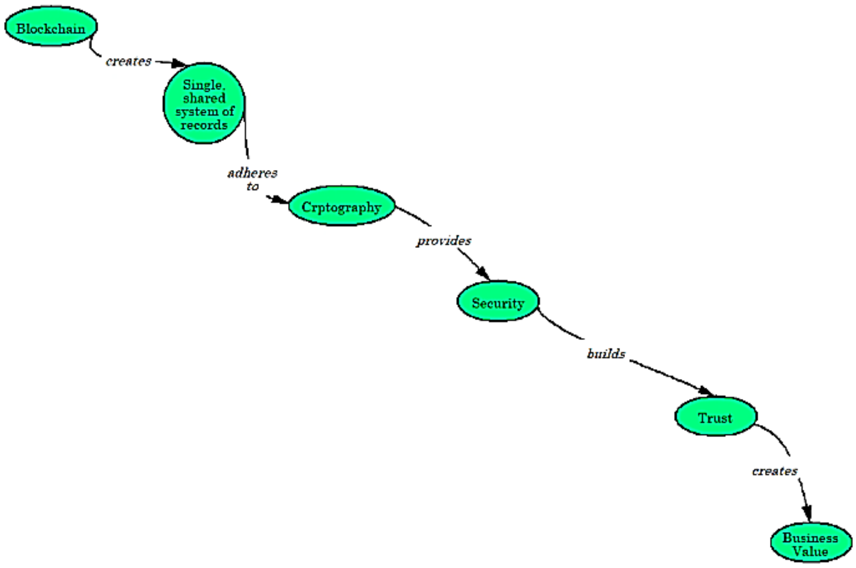


Figure 3: Mainstay of Blockchain systemigram

Figure 3 shows the mainstay of the blockchain systemigram, which results from a copious review of extant literature and its synthesis to identify the primary purpose of blockchain as a system. The findings suggest that the blockchain creates a single shared system of records, inter-organizational transactions expressed in financial terms. This system of records adheres to the principles of cryptography or encrypted transmission of transaction data, leading to secure transmission of data and protection of ownership rights (Min, 2019). The sense of security amongst the organizations builds trust, leading to business value creation through comprehensive linkage of high-volume transactions (Wang, Han, et al., 2019).

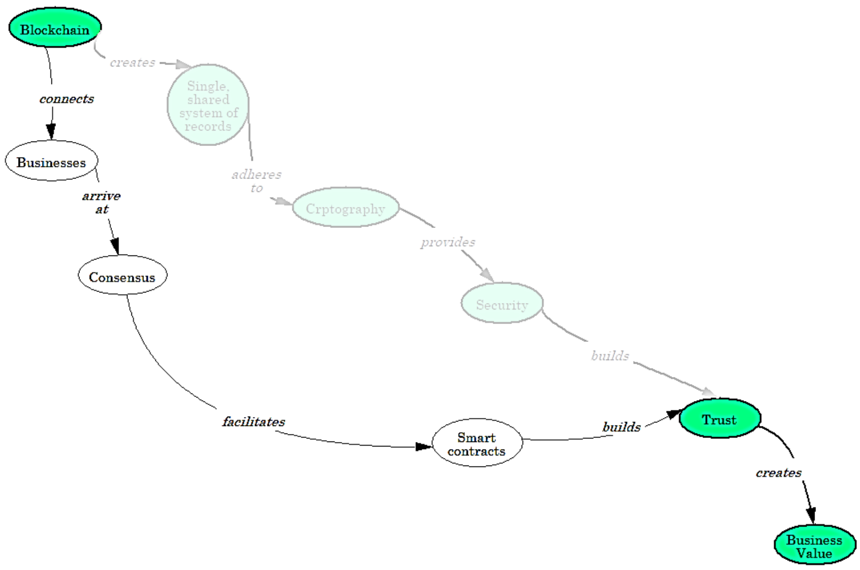


Figure 4: Blockchain systemigram scene 1

Figure 4 shows the role of blockchain in connecting multiple organizations in a single thread. Findings suggest that blockchain connects multiple businesses of organizations through functional and transactional linkages. These businesses or organizations arrive at a consensus, as facilitated by the functioning of the blockchain (Risius & Spohrer, 2017). The consensus further facilitates intelligent contracts, leading to building trust between the organizations and, ultimately, creating business value through linkages in transactions(Wang, Han, et al., 2019). A smart contract is an agreement with prescribed roles to govern business transactions. It is stored on the blockchain and automatically executed as part of the transaction (Wang, Han, et al., 2019). It contains contractual conditions to transfer resources between organizations. For example, when a flight is delayed by an agreed-upon duration, the smart contract, which contains travel insurance terms, is automatically triggered for execution. A smart contract is better than a traditional contract as it eliminates the hassles and delays in reaching a formal agreement by building the contract into inter-organizational transactions (Saberi, Kouhizadeh, Sarkis, & Shen, 2018). It establishes clear demarcated conditions under which transfer of assets or resources occurs, eliminating the need to transfer documents for multiple approvals (Min, 2019).

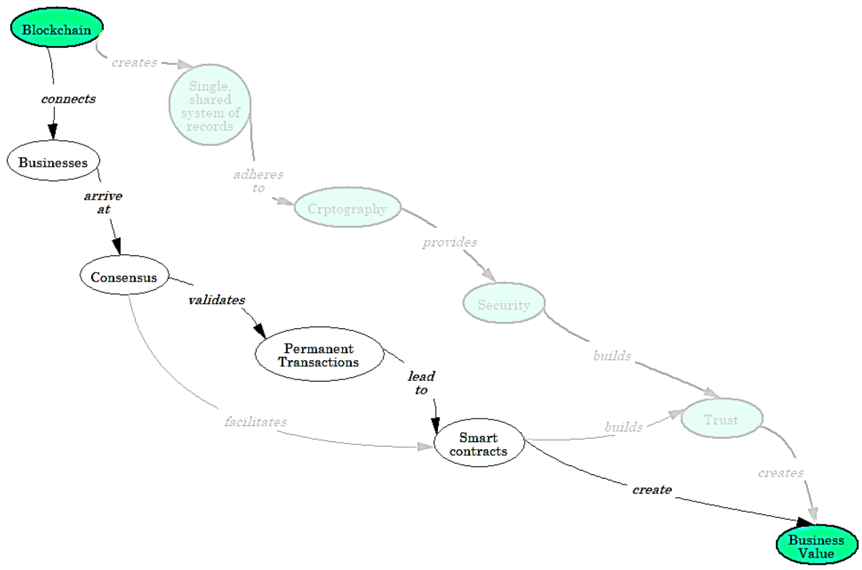


Figure 5: Blockchain systemigram scene 2

As shown in figure 5, the consensus reached amongst the organizations connected in the blockchain also validates the permanent transactions, leading to the creation of smart contracts (Saberi et al., 2018). This consensus results in creating business value for every organization in the blockchain. Consensus is built through algorithms that validate and authorize inter-organizational transactions (Kamble et al., 2018). These validated and authorized transactions are termed permanent transactions as they cannot be deleted or modified and exist for blockchain life unless all the organizations involved in the blockchain enter into another consensus mechanism to terminate the permanent transactions (Treiblmaier, 2018). This immutability allows the organizations participating in the blockchain to work at a speed synchronized with their business decision-making process (Felin & Wilson, 2018). The mechanism to reach consensus quickly is automated and varies across blockchains (Saberi et al., 2018). The mechanism of proof of stake is hinged on the notion that validation of transactions requires that the validating entity own a certain percentage of the inter-organizational network's complete validation (Risius & Spohrer, 2017). This proofing makes it very expensive to assume the role of a validator and prevents unauthorized ownership. The multi-signature mechanism entails consensus through transactional validation from multiple stakeholders (Queiroz & Fosso Wamba, 2019).

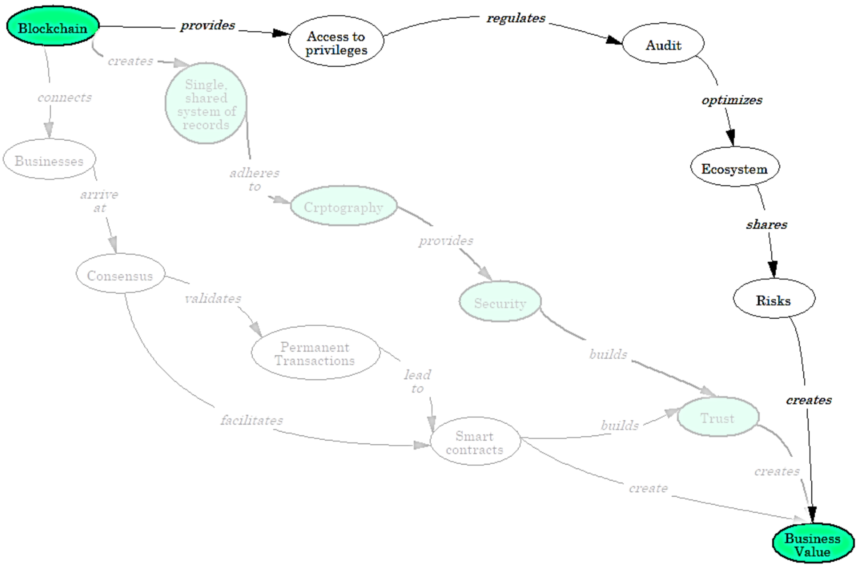


Figure 6: Blockchain systemigram scene 3

Figure 6 shows the significance of access, which organizations seek to leverage in accomplishing their objectives. Blockchain provides access to privileges, which may not have been otherwise available for organizations within its boundaries (Felin & Wilson, 2018). This access regulates the audit of the transactions conducted historically between organizations linked in the blockchain (Kshetri, 2018). The periodic audits optimize the ecosystem of the inter-organizational network, as organizations chart their further courses of action based on the audit results (Kshetri, 2018). The optimized ecosystem shares risks across multiple organizations and creates business value for the organizations in the blockchain (Felin & Wilson, 2018).

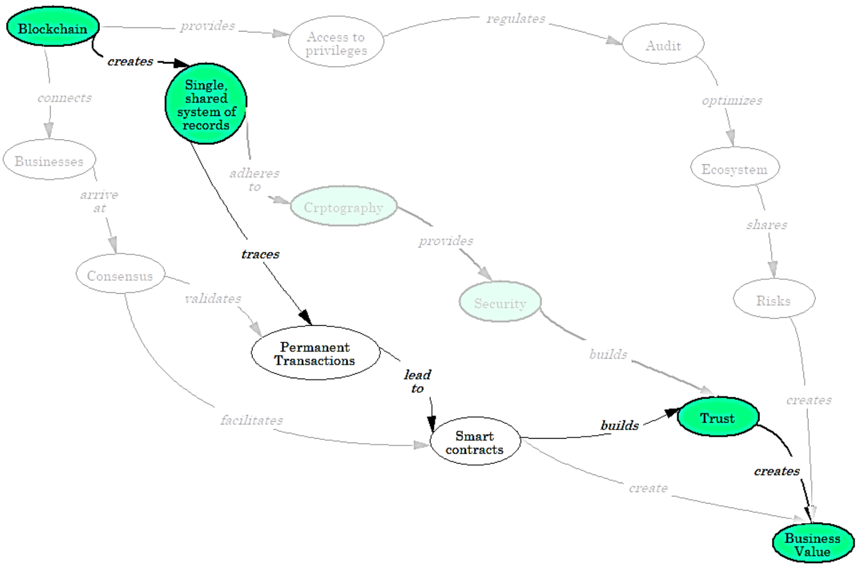


Figure 7: Blockchain systemigram scene 4

Figure 7 shows the role of a single shared system of records, as created by the blockchain. This system traces all the permanent transactions, leading to smart contracts (Wang, Singgih, et al., 2019). As discussed earlier, smart contracts build trust between the organizations, resulting in business value creation. Traditional accounting ledgers, based on bookkeeping principles, are in practice since medieval history. A shared and distributed ledger is advantageous over the traditional one because it is an immutable record of all the transactions on the blockchain network and is accessible to all the organizations in the blockchain (Risius & Spohrer, 2017). This sharing prevents duplication of transactional records but, at the same time, enables constituent organizations to get a copy for their record (Risius & Spohrer, 2017). The distributed ledger is regulated by levels of permissions, allowing only authorized access to view and share transactional information (Wang, Singgih, et al., 2019).

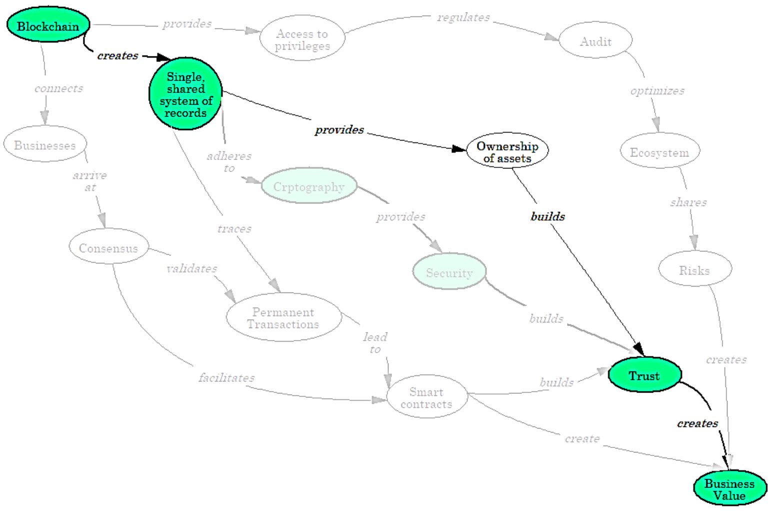


Figure 8: Blockchain systemigram scene 5

Figure 8 further shows the role of a single shared system of records or a distributed ledger. It provides rightful ownership of assets, leading to the building of inter-organizational trust and, ultimately, creating business value for the organizations in the blockchain (Felin & Wilson, 2018). The blockchain enables the transfer of ownership of assets between any two participating organizations in the blockchain network. This transfer is automatically recorded to the single shared system of records (Wang, Singgih, et al., 2019).

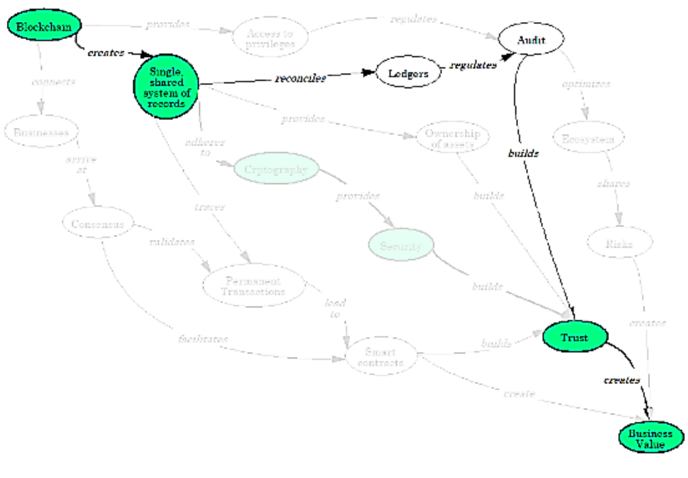


Figure 9: Blockchain systemigram scene 6

Figure 9 shows that the single shared system of records reconciles ledgers, or account-keeping documents, in a distributed ledger accessible to all the organizations connected in the blockchain. This reconciled ledger facilitates seamless, automated, and validated audits of the transactions, building trust amongst the organizations and creating business value (Wang, Singgih, et al., 2019). It also enables the blockchain to shift the paradigm from single owner to shared ownership of an asset or transaction (Felin & Wilson, 2018). Third-party intermediaries are eliminated as participating organizations can validate the transactions and verify identities based on roles and access privileges, pre-agreed at this stage (Kamble et al., 2018).

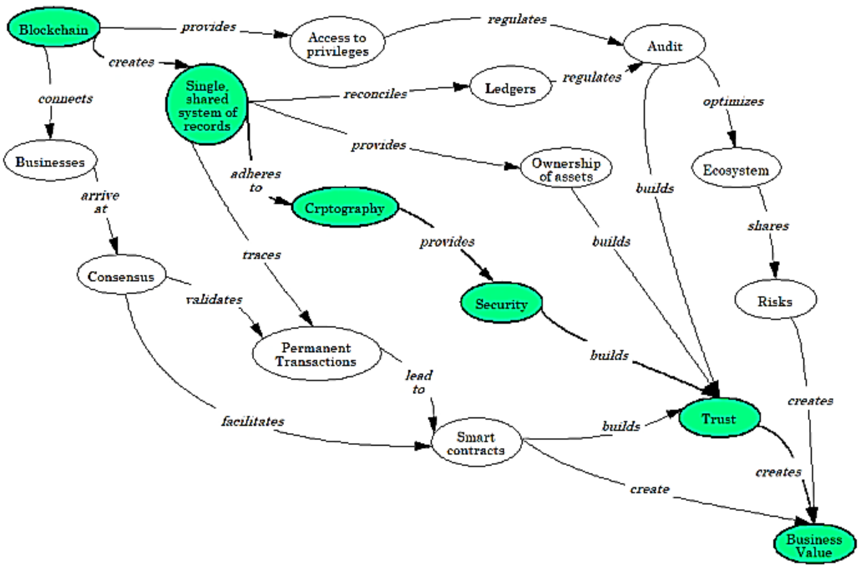


Figure 10: complete Blockchain systemigram

The comprehensive findings of applying systems thinking, based on Boardman Soft System Methodology, are shown in figure 10, a complete systemigram of the blockchain. It records multiple perceptions about the blockchain outcomes and organizational benefits in a single overarching perspective. Since the blockchain is a collection of organizations, which are controlled by people and have definitive driving objectives (Gorod, Sauser, & Boardman, 2012; Sauser et al., 2008), this study suggests that blockchain is a social system, as classified in the system thinking paradigm (Kembro et al., 2014).

DISCUSSION

The findings suggest the presence of whole system thinking concepts beneath the underlying mechanisms of functioning of blockchain. This section elaborates on the ways system thinking complements blockchain. The core aspects of blockchain, such as consensus, access, and sharing, are explored further through the analogy of systems thinking concepts for equilibrium (Sauser et al., 2008), common finality (Gorod et al., 2012), and self-regulation (Gandhi et al., 2012), respectively.

Equilibrium – value through consensus

Components in a system undergo fluctuations in their behavior so that the encompassing system reaches a steady or balanced state. When considering blockchain as a system, the constituent organizations are as components depicted in systems thinking. A steady or balanced state is defined as a state in which the components or organizations that drive the system's behavior as a whole are unchanging concerning time. This state represents the balance of matter, energy, and information flowing in, out, and across the system. A system in equilibrium resists or disregards the disturbance to defend itself against intrusion since a stable state can better predict the emerging behavior of the system. It uses homeostatic forces to restore the former balance if any component behaves sporadically to violate the equilibrium. As other components orient to cope with this behavior against the equilibrium, the system accommodates the disturbances by achieving a new equilibrium. This accommodation potentially explains the significance of consensus between organizations connected in a blockchain, which is crucial for creating business value through blockchain (Gandhi et al., 2012; Gonul Kochan et al., 2018; Wang, Han, et al., 2019).

Common finality – value through access

Although every component in the system has multiple objectives, a common goal is evident in the system's functioning. This common goal can be survival, power domination, etc. At the same time, this common goal is necessary for the system's survival. As per systems thinking, common finality refers to a typical or mutual goal that binds multiple components together despite their differences. It considers organizations or components as a set of parts that interact with each other, are organized, and manage to reach the same goal. The underlying assumption behind the concept of common finality is that every system struggles to search for a sustainable behavior capable of satisfying its operating context. Thus, in the context of a blockchain, organizations are managed by people to reach a common finality, which is necessary for terms of its environmental interactions. Smart contracts and access to privileges are the critical components of the blockchain system that capture mutual goals of the constituent organizations and sustain the inter-organizational bond to create a lasting business value through blockchain (John et al., 2018; Kshetri, 2018; Sauser & Boardman, 2015; Xi & Xia, 2017).

Self-Regulation – value through sharing

As per systems thinking, self-regulation is an adaptive mechanism that allows the system to keep itself under a balanced condition, within the limits of its structure, and through information exchange with the external world. The self-regulated systems can learn and perform a reflective knowledge function after interpreting the environment based on its longitudinal understanding or knowledge. The components of the system exhibit self-inhibiting or self-reinforcing behaviors because of the system's self-regulation, which is based on periodic self-evaluations in terms of accomplishing their objectives. These evaluations are further based on the feedback received by the components. Positive feedback leads to self-reinforcing behavior, while negative feedback leads to self-correcting or self-inhibiting behavior of the components. In blockchain functioning mechanisms, organizations seek to self-regulate themselves and others to achieve competitive advantages through harmony. The single shared system of records or distributed ledger is analogous to the self-regulating behavior of the blockchain as a system (John et al., 2018; Min, 2019; Sauser et al., 2008).

CONCLUSION

This study effectively demonstrates systems thinking, based on Boardman Soft System Methodology, using systemigram to conceptualize blockchain as a system. It combines multiple perspectives in a single narrative and offers future directions for empirical and interpretive research in multiple ways. First, the study suggests consensus, access to privileges, and distributed ledger in driving business value creation through blockchain. Future studies may further explore the role of these constructs empirically, using a network theory lens. Second, the potential of blockchain can be further explored through interpretive research to gauge whether the blockchain's hype is worth it. Third, this study offers insights to the organizations that seek to build blockchain and lead the inter-organizational network in a blockchain regarding organizational needs and their integration in an automated system.

REFERENCES –

Abeyratne, S. A., & Monfared, R. P. (2016). Blockchain Ready Manufacturing Supply Chain Using Distributed Ledger. *International Journal of Research in Engineering and Technology*, *05*(09), 1–10. https://doi.org/10.15623/ijret.2016.0509001

Chen, S., R. Shi, Z., Ren, J., Yan, Y. S., & Zhang, J. (2017). A blockchain-based supply chain quality management framework. *IEEE 14th International Conference on E-Business Engineering (ICEBE)*, 172–176.

Felin, T., & Wilson, C. E. (2018). What Problems Will You Solve With Blockchain? *MIT Sloan Management Review*, 7.

Gandhi, S. J., Gorod, A., & Sauser, B. (2012). A systemic approach to managing risks of SoS. *IEEE Aerospace and Electronic Systems Magazine*, *27*(5), 23–27. https://doi.org/10.1109/MAES.2012.6226691

Gonul Kochan, C., Nowicki, D. R., Sauser, B., & Randall, W. S. (2018). Impact of cloud-based information sharing on hospital supply chain performance: A system dynamics framework. *International Journal of Production Economics*, *195*(September 2016), 168–185. https://doi.org/10.1016/j.ijpe.2017.10.008

Gorod, A., Sauser, B., & Boardman, J. (2012). Computer Networks - 19th International Conference, Cn 2012, Szczyrk, Poland, June 19-23, 2012. Proceedings. *Journal of Computer Networks*, *2*(4), 50–61. https://doi.org/10.1007/978-3-642-31217-5_30

John, L., John, G. B., Parker, M., Sauser, B. J., & Wade, J. P. (2018). Self-organizing cooperative dynamics in government extended enterprises. *IEEE Systems Journal*, *12*(3), 2905–2916. https://doi.org/10.1109/JSYST.2016.2647598

Kamble, S., Gunasekaran, A., & Arha, H. (2018). Understanding the Blockchain technology adoption in supply chains-Indian context. *International Journal of Production Research*, *7543*. https://doi.org/10.1080/00207543.2018.1518610

Kembro, J., Selviaridis, K., & Näslund, D. (2014). Theoretical perspectives on information sharing in supply chains: A systematic literature review and conceptual framework. *Supply Chain Management*, *19*(April), 609–625. https://doi.org/10.1108/SCM-12-2013-0460

Korpela, K., Hallikas, J., & Dahlberg, T. (2017). Digital Supply Chain Transformation toward Blockchain Integration. *50th Hawaii International Conference on System Sciences*, 4182–4191. https://doi.org/10.1016/j.neuroimage.2018.01.059

Kshetri, N. (2018). 1 Blockchain's roles in meeting key supply chain management objectives. *International Journal of Information Management*, *39*(December 2017), 80–89. https://doi.org/10.1016/j.ijinfomgt.2017.12.005

Mandart, S. (2017). The Crypto Guide for Beginners – What is blockchain? Retrieved May 9, 2019, from CONDA AG, Donau-City-Straße 6, 1220 Vienna / Austria website: https://ico.conda.online/the-crypto-guide-for-beginners-—-what-is-blockchain/

Meadows, D. H. (2008). *Thinking in Systems - A Primer - Donella H. M.pdf* (6th ed.; D. Wright, Ed.). London: Earthscan.

Min, H. (2019). Blockchain technology for enhancing supply chain resilience. *Business Horizons*, *62*(1), 35–45. https://doi.org/10.1016/j.bushor.2018.08.012

Nakasumi, M. (2017). Information Sharing for Supply Chain Management Based on BlockChain Technology. *IEEE 19th Conference on Business Informatics (CBI)*, 140–149.

Omran, Y., Henke, M., Heines, R., & Hofmann, E. (2017). Blockchain-driven supply chain finance: Towards a conceptual framework from a buyer perspective. *(Working Paper)*, 15. Retrieved from https://www.alexandria.unisg.ch/251095/

Prins, R., Farr, J., McDonald, K., Fitzgerald, S., & Sanchez, D. (2015). Using Systemigrams and Fuzzy Cognitive Maps to Understand and Quantify Causality. *Proceedings of the 2015 Industrial and Systems Engineering Research Conference*, 193. Retrieved from http://uopx.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwtV3fS8MwED7mfBHEH6g4nRDweThd01wfROZo8UFFQcG3kTYp7KWttoW5v95c2mhB3JvPCUeO5C6X3PfdATCNXoAYJ7GIUZnwVoux9rWfItGBhSSScHh_9RjyaSiiHrhsbru1ziNaN63yhH7IL6iIZoDBJeJN8T6inIGUW3UNNP4WQAhM23NBXXscN2CTeKVkCO

Queiroz, M. M., & Fosso Wamba, S. (2019). Blockchain adoption challenges in supply chain: An empirical investigation of the main drivers in India and the USA. *International Journal of Information Management*, *46*(November 2018), 70–82. https://doi.org/10.1016/j.ijinfomgt.2018.11.021

Risius, M., & Spohrer, K. (2017). A Blockchain Research Framework. *Business & Information Systems Engineering*, *59*(6), 385–409. https://doi.org/10.1007/s12599-017-0506-0

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2018). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research*, *7543*. https://doi.org/10.1080/00207543.2018.1533261

Sauser, B., & Boardman, J. (2015). *Modeling and Simulation Support for System of Systems Engineering Applications* (1st ed.; L. Rainey & A. Tolk, Eds.). https://doi.org/10.1002/9781118501757.ch11

Sauser, B., Boardman, J., & Gorod, A. (2008). System of Systems Management. *System of Systems Engineering: Innovations for the 21*, *20*(4), 191–217. https://doi.org/10.1002/9780470403501.ch8

Sauser, B., Mansouri, M., & Omer, M. (2011). Using Systemigrams in Problem Definition: A Case Study in Maritime Resilience for Homeland Security. *Journal of Homeland Security and Emergency Management*, *8*(1). https://doi.org/10.2202/1547-7355.1773

Tapscott, D., & Tapscott, A. (2017). Will Change Organizations. *MIT Sloan Management Review*, *58*(2).

Tian, F. (2017). A Supply Chain Traceability System for Food Safety based on HACCP, Blockchain & Internet of Things. *International Conference on Service Systems and Service Management (ICSSSM)*, 1–6.

Toyoda, K., Takis Mathiopoulos, P., Sasase, I., & Ohtsuki, T. (2017). A Novel Blockchain-Based Product Ownership Management System (POMS) for Anti-Counterfeits in the Post Supply Chain. *IEEE Access*, *5*, 17465–17477. https://doi.org/10.1109/ACCESS.2017.2720760

Treiblmaier, H. (2018). The impact of the blockchain on the supply chain: a theory-based research framework and a call for action. *Supply Chain Management*, *23*(6), 545–559. https://doi.org/10.1108/SCM-01-2018-0029

Wang, Y., Han, J. H., & Beynon-Davies, P. (2019). Understanding blockchain technology for future supply chains: a systematic literature review and research agenda. *Supply Chain Management*, *24*(1), 62–84. https://doi.org/10.1108/SCM-03-2018-0148

Wang, Y., Singgih, M., Wang, J., & Rit, M. (2019). Making sense of blockchain technology: How will it transform supply chains? *International Journal of Production Economics*, *211*(November 2018), 221–236. https://doi.org/10.1016/j.ijpe.2019.02.002

Wu, H., Li, Z., King, B., Miled, Z. Ben, Wassick, J., & Tazelaar, J. (2017). A distributed ledger for supply chain physical distribution visibility. *Information (Switzerland)*, *8*(4), 1–18. https://doi.org/10.3390/info8040137

Xi, X., & Xia, Y. Q. (2017). An investigation of thermal and tribological behaviors of PTFE-based silicon composites filled with AlN and flake graphite particles. In *Journal of Applied Polymer Science* (Vol. 134). https://doi.org/10.1002/app.45263

DECISION SCIENCES INSTITUTE
Does Telehealth Reduce Rural-Urban Care-Access Disparities? Evidence from COVID-19 Telehealth Expansion

Shujing Sun, Guihua Wang *
Jindal School of Management, University of Texas at Dallas
shujing.sun@utdallas.edu, guihua.wang@utdallas.edu

ABSTRACT

We study the effect of telehealth expansion on rural-urban care-access disparities during COVID-19. We use urban areas as the control group and compare the changes in patients' access to care before and after the telehealth expansion. We distinguish in-person from telehealth visits and find enlarged disparities in patients' visiting modalities. In particular, urban patients substitute in-person visits with telehealth visits, yet rural patients have a much lower adoption rate of telehealth services and continued with in-person visits. Finally, we perform visit-level analyses and identify patients' social determinants and physicians' characteristics associated with telehealth adoptions.

KEYWORDS: COVID-19, Empirical health care, Rural-urban disparities, Telehealth, Expansion

INTRODUCTION

Rural areas of the U.S. have about 20% of the population but less than 10% of physicians.[1] The average travel distance from a patient to a physician is 10.5 miles in rural areas compared with 4.4 miles in urban areas.[2] The combination of a low physician-patient ratio and long travel distance has impeded rural patients' access to health care and, consequently, led to poor health outcomes. Studies have found rural Americans are worse in health conditions than their urban counterparts and more likely to die from heart diseases, cancers, and chronic lower respiratory diseases.[3] To reduce the rural-urban disparities in care access, the federal and state governments have implemented various policies to improve health care access in rural areas. For example, many states have expanded Medicaid to low-income adults. These expansions are critical to rural hospitals that are on the verge of closing and heavily relying on Medicaid reimbursements for health care services. To ease the shortage of physicians, a number of states have established legislation that expands the roles and responsibilities of non-physician primary care providers – such as nurse practitioners and physician assistants – when they practice in medically underserved communities such as rural areas.[4]

With advances in telehealth platforms and remote patient monitoring technologies, researchers and policymakers proposed using telehealth to reduce disparities in care between rural and urban areas. Telehealth has the potential to improve rural Americans' access to care for two reasons. First, it significantly reduces or even eliminates traveling costs, so patients can visit health care providers whenever needed. Second, patients have more options when choosing physicians, as telehealth allows them to visit faraway health care providers via virtual technologies. Despite its

* Authors contributed equally, and authors' names are shown in alphabetical order.

[1] See details at https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html

[2] https://www.pewresearch.org/fact-tank/2018/12/12/how-far-americans-live-from-the-closest-hospital-differs-by-community-type/

[3] https://www.cdc.gov/ruralhealth/about.html

[4] https://www.ncsl.org/research/health/improving-rural-health-state-policy-options-for-increasing-access-to-care.aspx

great promise, several barriers limit the extensive use of telehealth. First, physicians and patients were reluctant to use telehealth due to lower reimbursement rates, difficulty in establishing patient-provider relationships, and legal issues such as state licensure laws (Dorsey and Topol 2016). Second, according to CDC, patients of rural areas in the United States tend to be older than their urban counterparts.[5] Consequently, rural patients may be unwilling to use telehealth due to a lack of digital literacy. Besides, telehealth requires high-speed internet access and smartphones or related technologies. Rural Americans without these assets will not be able to use telehealth even if they are willing to visit health care providers virtually. Probably because of these barriers, the percentage of health care providers adopting telehealth is only 22% in 2019.[6]

While the use of telehealth was limited before 2020, the outbreak of COVID-19 has significantly propelled the adoption of telehealth services. The pandemic has blocked patients' access to their primary care physicians and hospitals. For those who manage to have an in-person visit, the exposure may significantly increase the risks of COVID-19 infection for both the patients and their health care providers. Telehealth can help reduce the impact of COVID-19 while ensuring patients having access to care. First, because telehealth allows patients to visit health care providers remotely, it reduces person-to-person contacts and the spread of the virus (Hollander and Carr 2020). Second, because telehealth minimizes the inconvenience of traveling, it allows patients to visit faraway health care providers when nearby providers face capacity issues. Recognizing the great potential of telehealth, policymakers called for an increase in access to telehealth services. Following the initiative, many payers had expanded the coverage of telehealth visits. For example, the Blue Cross and Blue Shield extended the coverage for telehealth services on March 19, 2020, by waiving cost-sharing for telehealth services for fully insured members and advocating for physicians' and health systems' adoption of this social distancing-encouraged capabilities.[7]

Despite the clear benefits of telehealth in containing the COVID-19 pandemic, it is unclear how the telehealth expansion affects care access and whether such effects might differ by rural and urban areas. More specifically, we aim to answer three important questions: (1) Does the expansion reduce the disparity of total care access in rural and urban areas? (2) How does the expansion change patients' visiting modalities (i.e., in-person or telehealth visits) in rural and urban areas? (3) What are some of the barriers that prevent health care providers and patients from adopting telehealth? Towards this goal, we collect national claims data of a large private health insurance provider that implemented the telehealth expansion policy. This payer expanded its telehealth coverage in March 2020. Before the telehealth expansion, patients have limited reimbursement for telehealth services. After the telehealth expansion policy, all members across the U.S. have access to an expanded list of procedures and services that are reimbursable when delivered via telehealth.

The main challenge to addressing these questions is that the telehealth expansion applies to patients of all areas around the same time. As a result, we cannot use a difference-in-differences (DID) approach to compare areas that did not expand telehealth with those that did to analyze the treatment effect. We address this challenge by using urban counties as the "control" group and compare the differential treatment effects between rural and urban counties before and after the expansion. This approach has been used in several other studies. For example, Cui et al. (2021) used male researchers as a control group to study the effect of COVID-19 (which applies to both male and female researchers) on gender inequality in research productivity. Another challenge is that telehealth expansion was introduced during COVID-19. Any changes in disparities between care in rural and urban areas may be due to COVID-19 instead of telehealth expansion.

[5] https://www.cdc.gov/ruralhealth/about.html

[6] https://www.beckershospitalreview.com

[7] https://www.bcbs.com/press-releases/media-statement-blue-cross-and-blue-shield-companies-announce-coverage-of-telehealth-services-for-members

To alleviate this concern, we conduct analyses with and without controlling for the COVID-19 infections and compare whether the results from these two models are significantly different from each other. The third challenge is that states implemented several other policies to contain the COVID-19 pandemic. For example, some states implemented stay-at-home orders, requiring residents to stay at home except for essential trips. A few other states issued guidance to suspend non-essential surgery in an effort to save limited health care resources for COVID-19 patients. There might be a concern that changes in disparities may be driven by these statewide policies. We address this concern by controlling a battery of COVID-19 related policies that affect patients' way of accessing care and were implemented during our sample period.

The findings suggest that rural-urban disparities in patients' overall access to care have not been much affected by the telehealth expansion policy. Interestingly, the composition of patients' visiting modalities changed dramatically due to the telehealth expansion policy. In particular, while the total number of visits from rural and urban areas follow a similar trend, urban patients' in-person visits are significantly reduced yet rural patients' in-person visits are not much affected. In other words, urban patients substitute in-person visits with telehealth visits. In contrast, rural patients have a much lower adoption rate of telehealth services and continued with in-person visits. Given that telehealth is preferred over in-person visits during the pandemic, we conclude that the disparities in the way of accessing care significantly increased. We also supplement the paper with visit-level analyses and identify patients' social determinants and physicians' characteristics that might have affected telehealth adoptions. As telehealth is likely to become an integral part of healthcare delivery, we believe our findings are important to researchers and providers in understanding the implications of telehealth expansion on rural-urban disparities and the existing barriers when expanding telehealth in rural areas.

LITERATURE REVIEW

Our study relates to two streams of literature: (1) use of telehealth, and (2) rural-urban disparity. We briefly review these two streams of literature in this section.[8]

Use of telehealth

Several studies have investigated the applications and impacts of telehealth in clinical settings. From the perspective of patients, Dorsey et al. (2013) conducted a randomized controlled clinical trial on 20 patients with Parkinson's disease. The paper finds web-based videoconferencing to provide specialty care at home is feasible, provides value to patients, and may offer clinical benefits similar to that of in-person care. Bavafa et al. (2018) studied the impact of e-visits on visit frequencies and patient health and found e-visit adoption increases the number of office visits and reduces the acceptance of new patients. Delana et al. (2019) studied the impact of telehealth centers in Southern India and found the opening of telehealth centers increases the overall network visit rate but reduces hospital visits, which suggests patients substitute hospital visits with telehealth center visits.

From the perspective of hospital operation and physician productivity, Saghafian et al. (2018) developed a partially observable Markov process to study the effectiveness of telemedical physician triage in workload management. The authors found lower-level agents should make decisions on a higher proportion of cases as the workload at the upper level increases. Çakıcı and Mills (2020) developed a Markov decision process to study the effectiveness of teletriage in managing healthcare demand. The authors found the addition of teletriage increases the rate of arrivals to the emergency department. Rajan et al. (2019) developed a game-theoretic model to compare the

[8] The first version of this paper was completed on February 18, 2021. To provide a complete picture of the current literature, we include papers written both before and after the completion of this paper. However, it is worth noting that some of the papers included in this literature review are incomplete working papers with preliminary results that are subject to change.

strategic behavior of revenue-maximizing and welfare-maximizing specialists. The authors found the latter serve a large patient population, spend less time with patients, and have shorter waiting times. Sun et al. (2020) studied the impact of telemedicine on emergency room congestion. The authors found telehealth availability significantly reduces emergency patients' waiting time and length of stay through flexible resource allocation.

Rural-urban disparity

The rural-urban disparity in healthcare access is a prevalent issue that has drawn increasing attention from researchers worldwide. Most of the existing studies compared the health of residents and access to providers in the U.S. For example, Cohen et al. (2018) studied associations between place-based characteristics of rural-urban status and health among adults above 65 and found an increase in urbanicity correlates with a decrease in negative self-reported health. Johnson et al. (2006) studied disparities in healthcare-provider availability in rural versus urban Alaska and New Mexico. The authors found rural residents have significantly less access to healthcare providers, and discrepancies increase with the level of required provider education and specialization. A number of studies examined health disparities in other countries and obtained similar results. For example, Liu et al. (2007) analyzed patterns in physician and hospital utilization among rural and urban populations in China and found rural residents use physicians more and hospitals less than urban residents. Masuda et al. (2018) studied regional differences between rural and urban areas in the management of acute myocardial infarction (AMI) in Japan and found AMI patients in rural areas are less likely to be transported directly to facilities with percutaneous coronary interventions.

With the advancement in information technology, researchers have found the potential of technology-mediated healthcare platforms and systems in alleviating the healthcare gap. For instance, Goh et al. (2016) studied the social value of an online health community and found urban users are net suppliers of social support, whereas rural participants are net recipients, which suggests online health communities can help alleviate rural-urban health disparities. Lyerly et al. (2016) studied the effect of telemedicine on access to acute stroke care for racial and ethnic minorities in Texas. They found telemedicine increases access to stroke expertise, and did not find evidence of disparities in access to the acute stroke expertise via telemedicine. Khairat et al. (2019) studied the effect of a virtual urgent care program on health equity and found the program reduces the disparity between rural and urban residents. Chunara et al. (2021) studied the use of telemedicine in a large healthcare system in New York City and found black patients are less likely to use telemedicine than white patients.

Compared with the existing literature, our paper has two notable contributions. First, the literature has recognized the potential of telehealth, yet the adoption of telehealth has not reached its full extent due to a lack of knowledge of its general implications. Using the telehealth expansion and the subsequent spike in telehealth adoption during the COVID-19 pandemic, we examine the causal impact of telehealth expansion on rural-urban disparity. Second, taking advantage of the unique large-scale data with individuals' visit modality (i.e., telehealth vs. in-person visit), we identify potential barriers to telehealth adoption and the effect of telehealth adoption on reducing COVID-19 infection. Our findings are critical to researchers and policymakers, because telehealth is likely to exist beyond the pandemic and become an integral part of healthcare delivery.

RESEARCH BACKGROUND AND DATA

In this section, we first describe the background of telehealth expansion. We then describe the datasets used in this study and discuss how we prepare them for empirical analyses.

Research background

The global spread of COVID-19 is an unprecedented and unexpected shock to the world. Following the national emergency declaration on March 13, 2020,[9] the government, providers, payers, and healthcare researchers have exerted relentless efforts to combat the COVID-19 outbreak. Among various initiatives, telehealth expansion has become a centerpiece that ensures patients' access to care while reducing the risks of COVID-19 exposure. For example, the Blue Cross Blue Shield Association, which provides health insurance to more than 106 million people in the United States, announced the coverage of telehealth services for its members on March 19, 2020.[10] Our paper aims to examine the impact of such telehealth expansion policy on patients' access to care and rural-urban care access disparities.

Data description

Our primary data[11] include individual patient claims to an anonymous payer from October 2019 to July 2020. This anonymous payer is one of the largest private health insurance providers, serving over 100 million people across the U.S. For each claim, we have information on the date of service, patient's demographic information (e.g., age, gender, and race), patient's diagnosis code, Current Procedural Terminology (CPT) code, and Healthcare Common Procedure Coding System (HCPCS) code, and the information of physicians who provided the care.

This payer expanded its telehealth coverage in March 2020. Before the telehealth expansion policy, patients have limited reimbursement for telehealth services. After the telehealth expansion policy, the payer waived cost-sharing for all members across the U.S. over an expanded list of procedures and services that are delivered via telehealth by in-network providers. More specifically, the anonymous payer in our sample covers telehealth codes consistent with the permanent code lists from the CMS-specified list.[12] The list covers a wide range of services, ranging from real-time telehealth visits between providers and patients, patient-initiated virtual check-in, and e-visits between patients and their providers through an online patient portal.[13] The expanded list also covers services for various diagnoses, such as psychotherapy, ophthalmological services, care for chronic diseases (e.g., diabetes, cardiac rehabilitation, and hypertension), general health and behavioral intervention, etc.

As we aim to investigate how telehealth expansion affects patients' access to care and visit modality, we restrict the sample to visits with procedures and services that are eligible for reimbursement under the CMS list, because patients associate with these visits are the direct beneficiaries of the telehealth expansion policy. We distinguish in-person from telehealth visits based on the modifiers appended to the HCPCS or CPT code for each claim. Specifically, claims with one of the following modifiers (G0, GT, GQ, and 95) are furnished via telehealth.[14] The modifier G0 identifies telehealth services for diagnosis, evaluation, or treatment of symptoms of an acute stroke. The modifier GT is the most commonly used modifier for telehealth claims, which recognizes telehealth services via interactive audio and video telecommunications systems. The modifier GQ identifies asynchronous telehealth services. The modifier 95 is a fairly new modifier and is used only when billing to private payers to indicate that services were rendered via synchronous telecommunication.

[9] https://www.whitehouse.gov/presidential-actions/proclamation-declaring-national-emergency-concerning-novel-coronavirus-disease-covid-19-outbreak/

[10] https://www.bcbs.com/press-releases/media-statement-blue-cross-and-blue-shield-companies-announce-coverage-of-telehealth-services-for-members

[11] Our study builds on a unique dataset from COVID-19 Research Database, which provides the de-identified claims data covering patients across the U.S.

[12] https://www.cms.gov/Medicare/Medicare-General-Information/Telehealth/Telehealth-Codes

[13] https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet

[14] https://www.bcbstx.com/provider/pdf/cpcp033-telemed-telehealth-02122021.pdf

Table 1 reports the visit-level summary statistics in our study. We see our sample includes more than 7.7 million visits during our study period. Comparing the columns *Before Expansion* and *After Expansion*, we see patient demographics, provider experience, and the proportion of individual practitioners remain almost the same after the telehealth expansion. However, the percentage of telehealth visits increases dramatically in both rural and urban areas, indicating a significant boost of telehealth visits after the policy change.

Table 1 Visit-level Summary Statistics

Variable	Before Expansion		After Expansion	
	Mean	Observation	Mean	Observation
Patient: Age				
Rural	36.51	177,786	36.71	310,450
Urban	33.99	2,678,034	34.80	4,545,858
Patient: Above 65%				
Rural	5.13%	177,786	5.93%	310,450
Urban	4.01%	2,678,034	4.64%	4,545,858
Patient: Female%				
Rural	62.25%	177,786	63.18%	310,450
Urban	59.37%	2,678,034	60.68%	4,545,858
Provider: Experience				
Rural	108.61	177,748	108.36	310,450
Urban	107.66	2,677,345	109.42	4,545,365
Provider: Individual Practitioner%				
Rural	38.26%	177,748	35.95%	310,450
Urban	33.27%	2,677,345	33.92%	4,545,365
Telehealth Visits%				
Rural	0.63%	177,786	23.60%	310,450
Urban	0.49%	2,678,034	37.55%	4,545,858

Note: This table reports the visit-level (i.e., claim-level) summary statistics of the main data. Provider experience is measured as the number of months since her/his registry of National Provider Identifier.

To account for confounding factors related to the COVID-19 situation and policies, we supplement the main data with the county-level COVID-19 cases from the Johns Hopkins University[15] and the state-level policies from the Kaiser Family Foundation (KFF).[16] In the propensity score matching (PSM), we further collect county-level characteristics from the 2018 Centers for Disease Control (CDC) Social Vulnerability Index, which contains 2014-2018 American Community Survey (ACS) estimates for various demographic and socioeconomic variables.[17]

ECONOMETRIC MODEL

In this section, we first illustrate the empirical challenges and our identification strategy. We then describe the DID model.

[15] https://coronavirus.jhu.edu/us-map

[16] https://www.kff.org/report-section/state-covid-19-data-and-policy-actions-policy-actions/

[17] https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2018.html

Empirical strategy

The main challenge we face in analyzing the impact of telehealth expansion is that all counties across the U.S. experienced the same expansion simultaneously, which makes directly applying the DID approach infeasible. Following the existing literature (Seamans and Zhu 2014, Cui et al. 2021), we extend the DID framework by using rural areas as the treatment group and urban areas as the control group. We then estimate the treatment effects by comparing the differences between rural and urban areas before and after the telehealth expansion.

The validity of this approach depends on two assumptions. The first assumption is the parallel-trends assumption, which is the most critical assumption. It requires that the outcome difference between the rural and urban areas stays the same in the absence of the expansion. The second assumption is that the expansion is exogenous to patients and providers. In other words, they did not anticipate the expansion, and hence they did not act strategically in deciding the modality of their clinical visits before the actual telehealth expansion.

To test whether the first assumption is violated, we follow the literature (Cui et al. 2021) by including pre-treatment dummies as independent variables. The second assumption cannot be tested empirically. However, because COVID-19 was an unexpected shock and the payer in our sample is one of the first to respond to the pandemic by delivering its telehealth expansion initiative, patients and providers were unlikely to foresee the policy change and adjust their behaviors ex-ante.

Difference-in-Differences model

We arrange the data at the county-month level and the specification is as follows:

$$Y_{i,t} = \alpha_0 + \alpha_1 Rural_i \times TeleExpansion_t + \beta X_{it} + CountyFE_i + TimeFE_t + \varepsilon_{it}, \qquad (1)$$

where $Y_{i,t}$ corresponds to the log-transformed number of hospital visits in county i at month t. We log-transform the dependent variable due to its skewed distribution. The transformation also facilitates the interpretation and comparison of coefficient estimates across different counties.[18] Since our first research question is on health care access disparity, we construct the first dependent variable as $ln(\textit{TotalVisits}_{i,t})$, which represents the log-transformed total number of visits in county i at month t. To further examine changes in patients' visiting modalities, we then construct $ln(\textit{In-personVisits}_{i,t})$ and $ln(\textit{TeleVisits}_{i,t})$, which represents the log-transformed number of in-person and telehealth visits in county i at month t, respectively. Following the 2013 NCHS Urban-Rural Classification Scheme, we define a county as rural (i.e., $Rural_i = 1$) if it is classified as "micropolitan" or "non-core".[19] The binary variable $TeleExpansion_t$ equals 1 for the post-expansion period (i.e., starting March, 2020).

The coefficient estimate α_1 of the interaction term captures the treatment effect of telehealth expansion on rural areas relative to urban areas. Because we have accounted for the county and time fixed effects (denoted by $CountyFE_i$ and $TimeFE_t$, respectively), the main effect of $Rural_i$ is absorbed by the county fixed effects, and $TeleExpansion_t$ is absorbed by the time fixed effects. In two separate analyses, we further include time-varying characteristics, X_{it}, such as county-level COVID-19 cases and the state-level policies as additional control variables. Throughout the analyses, we cluster the standard errors by states to account for possible unobserved correlations in error terms within each state. Because a stable composition of treatment and control groups in the DID model is critical for obtaining unbiased estimates of the treatment effects (Lechner et al. 2016), we drop counties with 20% or more missing observations at the monthly level and expand the panel to be balanced. Our motivation is to keep as many valid observations as possible while excluding outliers (i.e., counties with no patient visits at the month level).

[18] The results remain qualitatively the same when we construct the dependent variable in its original scale. Results are available upon request.

[19] https://www.cdc.gov/nchs/data_access/urban_rural.htm

MAIN RESULTS

In this section, we first estimate the main effects of telehealth expansion on total visits. We then distinguish in-person from telehealth visits and examine the changes in patients' visiting modality.

Effect on total visits

Before presenting the DID results, we begin with checking the model-free summary statistics on overall access (i.e., total visits) by patients' residences. Table 2 compares the total number of visits in rural and urban areas before and after telehealth expansion. Before the expansion, the average number of visits is 70.34 in rural areas and 760.43 in urban areas, so the difference before the expansion is $-687.39 (= 73.04 - 760.43)$. After the expansion, the number of visits is 102.78 in rural areas and 1084.64 in urban areas, so the difference after the expansion is $-981.86 (= 102.78 - 1084.64)$. The model-free DID estimate is thus $-294.47 (= -981.86 - (-687.39))$, suggesting a relative decrease in the volume of total visits of rural patients relative to urban patients. Note that this model-free analysis does not control for county and time fixed effects or time-varying county characteristics.

Table 2 Comparison of Total Visits Before and After Telehealth Expansion

Variable	Before Expansion			After Expansion			After - Before
	Mean	SD	Observation	Mean	SD	Observation	
Total Visit							
Rural	73.04	188.42	3252	102.78	244.06	2168	29.74
Urban	760.43	2515.40	4716	1084.64	3332.71	3144	324.21
Rural - Urban	-687.39			-981.86			-294.47

Note: This table shows the mean and standard deviation (SD) of total visits in rural and urban areas before and after the telehealth expansion. The “Observation” column indicates the number of county-month-level observations in the data.

To formally analyze the effect of telehealth expansion on patients' total access to health care, we execute the DID regression model (1). The results are reported in Table 3. In column (1), we observe a negative but insignificant treatment effect of the telehealth expansion policy, implying that the total number of visits evolved in parallel patterns and remains similar for rural and urban areas.

Because telehealth was expanded shortly following the COVID-19 pandemic, a natural concern regarding the baseline result is that the enlarged disparity between rural and urban providers in total healthcare access could be driven by the COVID-19 situation instead of the telehealth expansion. To alleviate this concern, we account for the COVID-19 situation, measured by the infection rate and death rate (i.e., the number of new COVID-19 cases and deaths per 100,000 population in a given month and county). As shown in column (2) of Table 3, the results remain qualitatively the same, suggesting the findings in total care access is not driven by the difference in the incidence of COVID-19.

As discussed in the Introduction, another empirical challenge is that states implemented other policies during the pandemic. As a result, patients' visit modality could be driven by concurrent policy changes regarding COVID-19 healthcare provision. For example, Wang (2021) found stay-at-home orders effectively reduce residents' mobility. To account for such a possibility, we control for the state-level stay-at-home orders and report the result in column (3) of Table 3. We see the coefficient estimate remains small and insignificant. Overall, the findings indicate that rural-urban care access disparities in overall access is not affected. Hence, the answer to the first question in the Introduction is no. The telehealth expansion policy does not reduce rural-urban care access disparities in our sample.

Table 3 Effect of Telehealth Expansion on Total Visits

Variable	ln(TotalVisits)		
	(1)	(2)	(3)
Rural $\times$ TeleExpansion	-0.023	-0.015	-0.0004
	(0.026)	(0.026)	(0.026)
County & Time FE	Y	Y	Y
Covid Cases	N	Y	Y
State Policy	N	N	Y
Observations	13,280	13,280	13,280
R-Squared	0.433	0.433	0.437

Note: *** p $<$ 0.01, ** p $<$0.05, * p $<$ 0.1. This table reports the DID estimates in equation 1 using the county-month level data from October 2019 to July 2020. The dependent variable is the log-transformed volume of in-person visits. Standard errors in parentheses are clustered by state.

Change in visit modality

To address the second question regarding patients' choices of visiting modality, we now turn to analyze the change in the number of visits by modality. Table 4 reports the summary statistics of in-person visits and telehealth visits in rural and urban areas before and after telehealth expansion. Before the expansion, the average number of in-person visits is 71.67 in rural areas and 742.68 in urban areas, so the difference before the expansion is $-671.01(= 71.67 - 742.68)$. After the expansion, the number of in-person visits is 72.71 in rural areas and 587.42 in urban areas, so the difference after the expansion is $-514.71(= 72.71 - 587.42)$. The model-free DID estimate is thus $156.30(= -514.71 - (-671.01))$. In other words, urban patients significantly reduced their in-person visits following the telehealth expansion. In contrast, rural patients' in-person visits did not change much. Regarding telehealth visits, we observe an increase in both rural and urban areas. However, the magnitude of increase is much larger in urban areas, which results in the non-parametric DID estimate being $-450.78(= -467.71 - (-16.37))$ for telehealth visits.

Table 4 Comparison of In-person and Telehealth Visits Before and After Telehealth Expansion

Variable	Before Expansion			After Expansion			After - Before
	Mean	SD	Observation	Mean	SD	Observation	
In-person Visits							
Rural	71.67	186.72	3252	72.71	185.45	2168	1.04
Urban	742.68	2485.58	4716	587.42	1566.89	3144	-155.26
Rural - Urban	-671.01			-514.71			156.30
Telehealth Visits							
Rural	1.37	8.12	3252	30.07	95.54	2168	28.70
Urban	17.74	146.11	4716	497.22	1905.33	3144	479.48
Rural - Urban	-16.37			-467.15			-450.78

Note: This table shows the mean and standard deviation (SD) of patients' in-person and telehealth visits in rural and urban areas before and after the telehealth expansion. The "Observation" column indicates the number of county-month-level observations in the data.

Following the regression model (1), we replicate the DID analysis by constructing the dependent variables as the log-transformed number of in-person and telehealth visits, respectively. The results are reported in Table 5. From columns 1–3, we observe a significant and positive treatment effect on rural patients. This implies that patients in rural areas have significantly greater tendencies to have in-person visits relative to those of urban areas. In terms of magnitude, compared to urban counties, rural counties experienced a 13.3% increase (see column 3) in in-person visits. From columns 4–6, we observe a significantly negative effect on telehealth visits, suggesting that rural areas are less likely to adopt telehealth than urban areas. Take column 6 as an example, rural patients' volume of telehealth visits is 158.3% lower than that of urban patients after the policy change. Overall, the findings show that urban patients substitute in-person visits with telehealth visits, but rural patients continue with the in-person visits.

Table 5 Effect of Telehealth Expansion on In-person and Telehealth Visits

Variable	ln(In-personVisits)			ln(TeleVisits)		
	(1)	(2)	(3)	(4)	(5)	(6)
Rural × TeleExpansion	0.125***	0.126***	0.133***	-1.716***	-1.627***	-1.583***
	(0.042)	(0.042)	(0.041)	(0.143)	(0.143)	(0.141)
County & Month FE	Y	Y	Y	Y	Y	Y
Covid Cases	N	Y	Y	N	Y	Y
State Policy	N	N	Y	N	N	Y
Observations	13,280	13,280	13,280	13,280	13,280	13,280
R-Squared	0.3484	0.3492	0.3499	0.6457	0.6516	0.6552

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table reports the DID estimates using the county-month-level data from October 2019 to July 2020, where the dependent variable is the log-transformed in-person visits (columns (1)–(3)) and log-transformed telehealth visits (columns (4)–(6)), respectively. Standard errors in parentheses are clustered by state.

ROBUSTNESS CHECK

To further evaluate the validity of our main empirical findings, we start with a pre-trend analysis to test if there is any violation of the parallel trend assumption. We then conduct a set of tests to account for various other state-level policies related to COVID-19. We also perform falsification tests to address endogeneity concerns due to unobserved seasonality. We conclude with propensity score matching, which enables us to estimate the treatment effect using a sample of comparable rural and urban counties.

Pre-trend analysis

A fundamental identification assumption in the DID model is the parallel trend assumption. To evaluate whether there are any signs of violation of this assumption, we create a series of lag indicators for each pre-treatment period and two lead indicators for the time periods following the telehealth expansion. More specifically, we expand the DID regression model as follows:

$$Y_{i,t} = \alpha_0 + \sum_{\tau=1}^{5} \alpha_{t-\tau} Rural_i \times Time_{t-\tau} + \alpha_{t=0} Rural_i \times Time_{t=0} + \alpha_{t\geq 1} Rural_i \times Time_{t\geq 1}$$
$$+\beta X_{it} + CountyFE_i + TimeFE_t + \varepsilon_{it},$$

where Y can be $ln($*TotalVisits*$)$ or $ln($*In-personVisits*$)$.[20] The coefficient estimates (α_{t-5}, α_{t-4},..., α_{t-1}) capture the month-by-month pre-treatment differences between rural and urban areas. $\alpha_{t=0}$

[20] Note that we do not need to test the pre-trend of $ln($*Televisits*$)$ because telehealth visits were minimal (almost zero) for most counties before the telehealth expansion. In other words, the parallel trend automatically satisfies between rural and urban areas.

and $\alpha_{t\geq 1}$ stand for the DID effect in the month and following months of telehealth expansion, respectively. The results are reported in Table 6, where the baseline period is 5-month before the expansion (i.e., 2019 October). We find no evidence of pre-trend differences in total visits and in-person visits between rural and urban areas, because the coefficient estimates for all the lag indicators are small and insignificant. We can see that the divergence of in-person visits between rural and urban areas emerged only since the policy implementation month.

Table 6 Pre-trend Analysis

Variable	ln(TotalVisits) (1)	ln(In-personVisits) (2)
Rural $\times$ Time$_{t-4}$	0.036	0.037
	(0.027)	(0.027)
Rural $\times$ Time$_{t-3}$	0.026	0.025
	(0.025)	(0.026)
Rural $\times$ Time$_{t-2}$	0.029	0.031
	(0.034)	(0.034)
Rural $\times$ Time$_{t-1}$	0.015	0.015
	(0.041)	(0.041)
Rural $\times$ Time$_{t=0}$	0.053	0.090**
	(0.044)	(0.045)
Rural $\times$ Time$_{t\geq 1}$	0.027	0.166***
	(0.034)	(0.048)
County & Month FE	Y	Y
Covid Cases	Y	Y
State Policy	Y	Y
Observations	13,280	13,280
R-Squared	0.4367	0.3502

Note: *** p $<$ 0.01, ** p $<$0.05, * p $<$ 0.1. This table reports the results from the pre-trend analysis. We use the five-month period before the telehealth expansion (i.e., $t-5$) as the baseline. The dependent variable is the log-transformed number of total and in-person visits for each county at the monthly level. Standard errors in parentheses are clustered by state.

State-level policy

Besides the stay-at-home orders, some states also issued guidance to suspend non-essential surgery to preserve limited healthcare resources for COVID-19 patients. Despite that this initiative is to curb non-essential procedures, it has been shown to also hurt the availability of essential ones (Wang and Dai 2020). Similarly, other state policies, such as surgery resumption and state reopen, may also affect patients' visiting modalities directly or indirectly. To check if these state-level policies drive the heterogeneous changes in patients' visiting behaviors between rural and urban areas, we replicate the DID analysis (equation 1) with additional controls of three other state-level policies that were implemented during our sample period and might affect patients' modalities to access health care. In particular, besides the stay-at-home orders that were implemented around the same time as the telehealth expansion policy, we further control for policies regarding the suspension of elective medical procedures, the resumption of elective medical procedures, and the reopen.[21] Since the policy implementation date differs by state, we generate binary variables to indicate the post-policy periods of different policies based on a patient's location. Table 7 reports the regression results using different dependent variables. We find consistently robust results, suggesting that our findings are not driven by these confounding policy implementations.

Table 7 Control for Additional Covid-Related Policies

Variable	ln(TotalVisits) (1)	ln(In-personVisits) (2)	ln(TeleVisits) (3)
Rural × TeleExpansion	-0.001	0.131***	-1.582***
	(0.027)	(0.040)	(0.140)
Stay-at-home Order	0.181**	0.085	0.504**
	(0.071)	(0.091)	(0.235)
Surgery Suspension	-0.057	-0.035	-0.055
	(0.044)	(0.073)	(0.163)
Surgery Resumption	0.081***	0.042	0.052
	(0.023)	(0.048)	(0.102)
State Reopen	-0.030	0.065	-0.174
	(0.027)	(0.061)	(0.13)
County & Month FE	Y	Y	Y
Covid Cases	Y	Y	Y
State Policy	Y	Y	Y
Observations	13,280	13,280	13,280
R-Squared	0.437	0.350	0.655

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table reports the DID estimates with additional controls for other state-level COVID-19 policy. Besides stay-at-home order, we include state reopening, non-essential surgery suspension, and resumption policies during our study window. The dependent variable is the log-transformed monthly visits by modality for each county. Standard errors in parentheses are clustered by state.

Falsification test: Pseudo-treatment time

Another related concern is that the identified treatment effect could be due to unobserved seasonality in March. To alleviate this concern, we replicate the analysis in equation 1 by assuming the pseudo-treatment time to be March of 2019, one year ahead of the actual policy implementation. The rationale is as follows. If our findings were driven by unobserved seasonality of patient volume between rural and urban areas, we would be able to falsely detect a similar result in other years around the policy time. The falsification tests are reported in Table 8, which is based on the same set of counties' data from 2018 October to 2019 July (i.e., one year before our main sample period). We do not find any treatment effects in the placebo year for $ln(\textit{TotalVisits})$ and $ln(\textit{In-personVisits})$. We do find a significant coefficient estimate when the dependent variable is $ln(\textit{TeleVisits})$. The estimate in column (3) is significant but the direction of the effect is the opposite of the treatment effects in Table 5, thereby suggesting that our main findings are not driven by unobserved seasonality around the month March.

[21] https://www.mcguirewoods.com/client-resources/Alerts/2020/10/state-governors-stay-at-home-prohibition-elective-procedures-orders

Table 8 Falsification Test with Pseudo-treatment Time (i.e., 2019 March)

	ln(TotalVisits) (1)	ln(In-personVisits) (2)	ln(TeleVisits) (3)
Rural × TeleExpansion	-0.009	-0.007	0.051**
	(0.033)	(0.033)	(0.023)
County, Month FE	Y	Y	Y
Covid Cases	Y	Y	Y
State Policy	Y	Y	Y
Observations	13,280	13,280	13,280
R-Squared	0.473	0.472	0.027

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table reports the falsification tests by assuming the pseudo-treatment time to be March of 2019. The dependent variable is the log-transformed number of monthly visits by modality for each county. Standard errors in parentheses are clustered by state.

Propensity score matching

Despite a wide range of controls, one may still worry the finding could be driven by the selection bias between rural and urban patients. For example, rural areas typically have smaller patient volumes relative to care providers and therefore lower risks of infection due to in-person visits. Accordingly, rural patients may worry less about going to the doctor's office and have more in-person visits than urban patients. Consequently, we will observe a relative increase in rural patients' in-person visits, regardless of the telehealth expansion. To further alleviate the concern, we refer to PSM (Rosenbaum and Rubin 1983). The idea is to reduce selection bias by making the treatment and control groups (i.e., rural and urban counties) more comparable with respect to the pre-treatment characteristics.

Table 9 Impact of Telehealth Expansion on Hospital Visits by Modality - PSM

	ln(TotalVisits) (1)	ln(In-personVisits) (2)	ln(TeleVisits) (3)
Rural × TeleExpansion	0.009	0.153**	-0.497***
	(0.046)	(0.069)	(0.151)
County, Month FE	Y	Y	Y
Covid Cases	Y	Y	Y
State Policy	Y	Y	Y
Observations	7,080	7,080	7,080
R-Squared	0.346	0.276	0.449

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table reports the DID estimates in equation (1) using the matched sample of rural and urban counties. The dependent variable is the log-transformed number of monthly visits by modality for each county. Standard errors in parentheses are clustered by state.

To implement the PSM, we start with a cross-sectional sample of rural and urban counties, and match counties based on various characteristics *before* telehealth expansion. More specifically, we match by the 3-month historical visits (i.e., from 2019 July to 2019 September), so that the matched counties have similar patient volumes before the policy change. We also include a wide range of variables from the 2018 CDC Social Vulnerability Index in the matching, including healthcare access, socioeconomic characters, household senior and disability composition, minority status, and transportation condition. Follow the literature (Stuart 2010), we allow for replacement in the matching since there are much fewer control counties (i.e., urban areas) that are comparable to the treatment counties (i.e., rural areas) in variables such as patient volume and population size. Using propensity score matching with a caliper of 0.01 and allowing for replacement, we identify 708 matched counties (542 rural counties and 166 urban counties).[22] The matched sample is balanced in all matching variables. Table 9 reports the regression results. Regardless of the differences in model specifications, the results consistently support our main findings across all dependent variables.

DISCUSSION AND INSIGHTS

In this section, we first identify some of the barriers that prevent healthcare providers and patients from adopting telehealth. We then analyze whether telehealth adoption can help reduce Covid-19 infections.

Barriers to telehealth adoption

We now address the third question from the Introduction by performing visit-level analyses and identifying patients' social determinants and physicians' characteristics that might have affected their adoption of telehealth.

From the patient side, we consider how social determinants associate with one's telehealth adoption. We start with identifying patient visits to providers who adopted telehealth into their practice. For example, if a provider conducted her/his first telehealth visit on April 3, 2020, we define this provider as an adopter and include only clinical visits to this provider after April 3, 2020. The rationale is that given a provider is an adopter, the modality for a visit depends more on a patient's choice. Using the visit-level data in this subsample, we conduct a logistic regression and report the results in column (1) of Table 10. We see patients from rural areas, patients above 65, and male patients are less likely to use telehealth. Consistent with the telehealth initiative to overcome the geographic barriers, we observe patients who are in different locations from providers (i.e., patients and providers do not share the same zip code) are more likely to utilize telehealth.

From the provider side, we consider how the experience and organizational affiliation affect one's telehealth adoption. Similarly, we perform a logistic regression using visit-level data from patients who are telehealth adopters (i.e., utilized telehealth at least once). Within this sample, given a patient's willingness to use telehealth, we can pin down factors that affect providers' choice of visit modality. The results are shown in column (2) of Table 10. We find more experienced providers are more likely to adopt telehealth. Interestingly, physicians affiliated with an organization are less likely to adopt telehealth than individual practitioners, probably because the latter have more flexibility in setting their practice patterns. In both analyses, we also account for the clinical classification code of a patient's primary diagnosis, which may directly affect one's telehealth utilization.

Effect of telehealth adoption on COVID-19 infection

Besides ensuring patients get the health care they need, telehealth intends to promote the practice of social distancing. To examine whether telehealth adoption reduces COVID-19 infection, we refer to the individual data and investigate the correlation between telehealth adoption and the likelihood of COVID-19 infection. For the following analysis, we subset the visit-level data after the policy enactment, when telehealth became available to patients. We identify a patient's COVID-19

[22] Note that the result of propensity score matching is not sensitive to the threshold of calipers 0.05 or 0.1. Results corresponding to different parameter setups are available upon request.

Table 10 Barriers to Telehealth Adoption

Variable	Telehealth Adoption Patient (1)	 Provider (2)
Rural	-0.576*** (0.006)	
Age: Above 65	-0.152*** (0.007)	
Male	-0.395*** (0.003)	
Different Locations	0.116*** (0.003)	
lnExperience		0.050*** (0.003)
Entity Type - Organization		-0.368*** (0.005)
State & Month FE	Y	Y
Primary Diagnosis	Y	Y
COVID Cases	Y	Y
State Policy	Y	Y
Observations	3,227,719	2,024,997
Pseudo R-Squared	0.205	0.090

Note: *** p $<$ 0.01, ** p $<$0.05, * p $<$ 0.1. This table reports the logistic regression using the visit-level data, where the dependent variable is a binary variable indicating whether a visit is via telehealth. Column (1) corresponds to the sample of providers who adopted telehealth. Column (2) corresponds to the sample of patients who adopted telehealth. The variable, *Different Locations*, indicates whether a patient and a provider have the same zip code.

infection status using the source data on one's primary diagnosis code, where a diagnosis of U07. 1 or U07. 2 indicates an infection.

For an individual i at time t, we create two time series: (1) the visit modality of individual i at time t (denoted by $TelehealthVisit_{i,t} \in \{0, 1\}$); (2) whether the individual i is infected with Covid-19 at time t (denoted by $Infection_{i,t} \in \{0, 1\}$). For the kth day forward (i.e., day $t + k$), we then analyze all individual observations and calculate the average infection rate as the percentage of infected patients, conditional on the visit modality at day t (i.e., $Pr[Infection_{t+k} = 1|TelehealthVisit_t = 0]$ and $Pr[Infection_{t+k} = 1|TelehealthVisit_t = 1]$).

Table 11 reports the results, where we re-scale the infection rate by a factor of 1/10,000 to allow for more non-zero digits in the statistics. We can see patients who adopted telehealth indeed had a much lower infection rate than those who went through in-person visits. The t-test results suggest that the average infection rates are indeed significantly different between the two visit modalities. Moreover, the difference in the infection rate following these two visit modalities seems to become more salient over time. Although this is just a non-parametric analysis, and we are not arguing for a causal story, the results here at least suggest the potential of telehealth in preventing Covid-19 spread. As a result, we consider telehealth a preferred way of accessing care than in-person visits during the pandemic.

Table 11 Telehealth and COVID-19 Infection

Days Relative to Visit	Infection Rate			
	In-Person Visit	Telehealth Visit	Difference	t-statistic
2 days after visit	0.944	0.429	-0.515***	6.202
4 days after visit	1.544	0.835	-0.709***	6.533
6 days after visit	1.994	1.304	-0.690***	5.454
8 days after visit	2.407	1.642	-0.765***	5.467
10 days after visit	2.653	1.789	-0.864***	5.888
12 days after visit	2.905	2.009	-0.896***	5.815
14 days after visit	3.154	2.167	-0.987***	6.153

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. This table reports the average future infection rate conditional on patients' visit modality. We rescale the infection rate by a factor of 1/10000 to allow for more non-zero digits in the statistics.

CONCLUSION

Based on a unique dataset with detailed visit-level information, we evaluated the impact of telehealth policy on care access disparities between rural and urban areas. Contrary to the well-intention of extending patients' swift access to care during the pandemic, telehealth expansion disproportionately benefits urban patients more than rural patients. Robust empirical results suggest that the total patient visits evolved similarly in rural and urban areas, thereby rural-urban disparities in total access to care did not reduce due to telehealth expansion. Moreover, rural patients have 12.5% more office visits than urban patients following the telehealth expansion, yet 158.3% fewer telehealth visits than urban patients. We also find that the utilization of telehealth associates with a lower infection rate. Overall, the results imply that urban patients benefit from the expansion by having significantly more telehealth visits and fewer in-person visits, while rural patients continue with office visits and are at higher risks of infections. Through individual-level analysis, we further examine several barriers to telehealth adoption. We find that social determinants (i.e., geographic location, age, gender, travel distance, providers' organizational affiliation) significantly correlate with one's use of telehealth visits.

Our paper adds to the literature on rural-urban disparity, a significant and long-standing social issue, especially within the healthcare context. Our results also contribute to the nascent literature on health IT and telehealth in particular. Using the telehealth expansion as a shock, we identify how patients of different social-economic statuses reacted to the telehealth availability and how this affects the rural-urban disparity regarding the channels of access to care. Our findings indicate that, despite the overall increase in telehealth usage following the telehealth extension, rural patients are significantly disadvantaged than their urban counterparts and have much lower adoption rates of this relatively safe option in accessing care in the pandemic. We hope that our findings could increase the awareness of the issue. For healthcare decision-makers, they may consider taking further actions to clear the roadblocks to "truly" increase rural patients' access to telehealth. As the COVID-19 pandemic accelerated the trend toward remote health, policymakers shall take this opportunity, bear in mind the issue of rural-urban disparity, and carefully design the policy change.

References

Bavafa H, Hitt LM, Terwiesch C (2018) The impact of e-visits on visit frequencies and patient health: Evidence from primary care. *Management Science* 64(12):5461–5480.

Çakıcı ÖE, Mills AF (2020) On the role of teletriage in healthcare demand management. *Manufacturing & Service Operations Management* 0(0), https://pubsonline.informs.org/doi/10.1287/msom.2020.0908.

Chunara R, Zhao Y, Chen J, Lawrence K, Testa PA, Nov O, Mann DM (2021) Telemedicine and healthcare disparities: A cohort study in a large healthcare system in New York City during COVID-19. *Journal of the American Medical Informatics Association* 28(1):33–41.

Cohen SA, Cook SK, Sando TA, Sabik NJ (2018) What aspects of rural life contribute to rural-urban health disparities in older adults? Evidence from a national survey. *The Journal of Rural Health* 34(3):293–303.

Cui R, Ding H, Zhu F (2021) Gender inequality in research productivity during the COVID-19 pandemic. *Manufacturing & Service Operations Management* (Forthcoming).

Delana K, Deo S, Ramdas K, Babu G, Ravilla T (2019) Multichannel delivery in healthcare: The impact of telemedicine centers in Southern India. *Available at SSRN 3505318* http://dx.doi.org/10.2139/ssrn.3505318.

Dorsey ER, Topol EJ (2016) State of telehealth. *New England Journal of Medicine* 375(2):154–161.

Dorsey ER, Venkataraman V, Grana MJ, Bull MT, George BP, Boyd CM, Beck CA, Rajan B, Seidmann A, Biglan KM (2013) Randomized controlled clinical trial of “virtual house calls” for Parkinson disease. *JAMA Neurology* 70(5):565–570.

Goh JM, Gao G, Agarwal R (2016) The creation of social value: Can an online health community reduce rural–urban health disparities? *MIS Quarterly* 40(1).

Hollander JE, Carr BG (2020) Virtually perfect? Telemedicine for COVID-19. *New England Journal of Medicine* 382(18):1679–1681.

Johnson ME, Brems C, Warner TD, Roberts LW (2006) Rural–urban health care provider disparities in Alaska and New Mexico. *Administration and Policy in Mental Health and Mental Health Services Research* 33(4):504–507.

Khairat S, Haithcoat T, Liu S, Zaman T, Edson B, Gianforcaro R, Shyu CR (2019) Advancing health equity and access using telemedicine: A geospatial assessment. *Journal of the American Medical Informatics Association* 26(8-9):796–805.

Lechner M, Rodriguez-Planas N, Fernández Kranz D (2016) Difference-in-differences estimation by FE and OLS when there is panel non-response. *Journal of Applied Statistics* 43(11):2044–2052.

Liu M, Zhang Q, Lu M, Kwon CS, Quan H (2007) Rural and urban disparity in health services utilization in China. *Medical Care* 767–774.

Lyerly MJ, Wu TC, Mullen MT, Albright KC, Wolff C, Boehme AK, Branas CC, Grotta JC, Savitz SI, Carr BG (2016) The effects of telemedicine on racial and ethnic disparities in access to acute stroke care. *Journal of Telemedicine and Telecare* 22(2):114–120.

Masuda J, Kishi M, Kumagai N, Yamazaki T, Sakata K, Higuma T, Ogimoto A, Dohi K, Tanigawa T, Hanada H, et al. (2018) Rural-urban disparity in emergency care for acute myocardial infarction in Japan. *Circulation Journal* CJ–17.

Rajan B, Tezcan T, Seidmann A (2019) Service systems with heterogeneous customers: Investigating the effect of telemedicine on chronic care. *Management Science* 65(3):1236–1267.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Saghafian S, Hopp WJ, Iravani SM, Cheng Y, Diermeier D (2018) Workload management in telemedical physician triage and other knowledge-based service systems. *Management Science* 64(11):5180–5197.

Seamans R, Zhu F (2014) Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Science* 60(2):476–493.

Stuart EA (2010) Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 25(1):1.

Sun S, Lu SF, Rui H (2020) Does telemedicine reduce emergency room congestion? Evidence from New York State. *Information Systems Research* 31(3):972–986.

Wang G (2021) Using mobile device data to understand the effect of stay-at-home orders on residents' mobility. *Manufacturing & Service Operations Management* (Forthcoming).

Wang G, Dai T (2020) The spillover effect of suspending non-essential surgery: Evidence from kidney transplantation. *Available at SSRN 3719662* 0(0), http://dx.doi.org/10.2139/ssrn.3719662.

DECISION SCIENCES INSTITUTE

Short-term Prediction of ICU Admission for COVID-19 Inpatients

Yoon Sang Lee
Columbus State University
Email: yslee211@gmail.com

ABSTRACT

Since the COVID-19 outbreak, many hospitals suffered from a surge of some high-risk inpatients needing to be admitted to the ICU. In this study, we propose a method predicting the COVID-19 inpatients' admission to ICU within a time frame of 12 hours. Bayesian Ridge Regression-based missing value imputation, the synthesis of training samples by the combination of two rows of each patient, and XGBoost are used for the proposed method. In the experiment, the AUC-ROC of our method is compared with those of other methods using various imputation techniques and classifiers. Our method achieves the best performance among the methods.

KEYWORDS: Machine Learning, XGBoost, COVID-19, Missing Values

INTRODUCTION

Since the COVID-19 outbreak began in 2019, many hospitals suffered from a surge of high-risk inpatients needing to be admitted to intensive care units (ICU). However, not all such patients could be admitted to the ICU owing to the limited resources in ICUs. Aziz et al. (2020) recommended developing a mathematical model to predict the demand of ICUs for improved organizational management. As most COVID-19 inpatients might not have detailed medical records at the moment of hospital admission and there would not be enough time to collect the patient's medical information, a prediction method using limited sets of medical measurements would be beneficial not only for the prediction of COVID-19 inpatients' admission to ICU, but also for the patients in similar situations from other diseases. While long-term (more than 15 days) prediction models for the future medical events including the COVID-19-related symptom changes and the general medical status changes have been proposed, to the best of our knowledge, individual level models predicting the short-term ICU admission of a COVID-19 patient within several hours have not been developed. When the demand for resources in an ICU increases in a short time due to a sudden surge of patients who need to be admitted to the ICU, the preparation of the resource within the time limit would not be easy. Therefore, if we could predict the patients who would be admitted to ICU in a short-term period, the planning of resource allocation for ICUs could be more efficient. To fill the gap, we propose a method predicting COVID-19 inpatients' admission to ICU within a short-term period. In the method of this study, we use a three-step process to train the model. In the first step, we apply Bayesian Ridge Regression (BRR) to each feature of each patient to impute the missing values. In the second step, we synthesize the training samples by combining two rows of each patient. This step is conducted to increase the variability of training samples. As the last step, we train Extreme Gradient Boosting (XGBoost) by using the imputed training set. In the experiments, the Area Under the Curve-Receiver Operating Characteristics (AUC-ROC) is used to measure the performance and our method achieves the best performance (AUC-ROC = 0.823) among methods in the experiment. The result implies that the method proposed in this study (BRR imputation + synthesis of training samples using combinations + XGBoost) is an applicable method for the prediction of the likelihood of COVID-19 inpatients' admission to the ICU.

The rest of the paper is organized as follows: Section 2 presents the related works. In Section 3, the details of data used in the study are explained. The method of this study is described in Section 4. Section 5 includes the experiment setup, its result, and a discussion of the result. Section 6 presents the conclusion of this study.

LITERATURE REVIEW

Studies for COVID-19 Patients' ICU Admission

To predict the ICU demand for COVID-19 patients, several studies suggested machine learning (ML) models. Gomes et al. (2020) proposed a model identifying the patients who will be developing severe symptoms and admitted to ICU. This model uses X-ray images to extract features and analyzes the features using a decision tree. Roncon et al. (2020) also analyzed the risk of ICU admission and the mortality of diabetes patients. The result showed that patients having diabetes have a higher risk of ICU admission and death. He et al. (2020) found that patients having cardiovascular disease have a higher risk of developing severe symptoms, requiring ICU admission and respiratory support treatment. A model yielding a risk score for the ICU admission and the mortality of COVID-19 patients was proposed by Zhao et al. (2020). This study was a retrospective study using the data collected over approximately 1.5 months. The model proposed by Zhao et al. (2020) yielded the AUC-ROC of 0.74 for the prediction of ICU admission and 0.83 for the prediction of mortality.Relating to COVID-19, some studies proposed models predicting COVID-19 patients' symptom development. Gao et al. (2020) adopted an ensemble of four classifiers for the prediction of physiological deterioration and death up to 20 days. Those four classifiers are: logistic regression, support vector machines (SVM), gradient-boosted decision tree, and neural networks (NN). Sun et al. (2020) identified 36 clinical indicators relevant to severe/critical symptoms and developed a prediction model using SVM. Assaf et al. (2020) proposed a model combining machine learning techniques and APACHE II risk prediction score for the prediction of the COVID-19 patients with non-severe, severe, and critical status.

As the spreading of COVID-19 is fast, many studies focus on the short-term prediction. A group of studies proposed the models predicting the spread of COVID-19 and the demand of ICU in a short-term period (Bekker and Koole 2021; Berta et al. 2020; Bonnasse-Gahot et al. 2020; Català et al. 2020; Chin et al. 2020; El-Ghitany 2020; Farcomeni et al. 2021; Funk et al. 2020; Goic et al. 2021; Keeling et al. 2020; Manevski et al. 2020; Massonnaud et al. 2020; Petermann et al. 2021; Rahaman Khan and Hossain 2020; Ricoca Peixoto et al. 2020; Weissman et al. 2020; C. Zhao et al. 2020; Zhao et al. 2021). These studies focused on the status changes of a population in certain regions (e.g. the short-term prediction of ICU demand in European countries). Another group of studies adopted computer tomography (CT) for the short-term prediction of lung anomalies (Chassagnon et al. 2020), the CT score-based prognosis (Francone et al. 2020), and ventilation (Burdick et al. 2020) of COVID-19 patients. Some studies tracked the changes of medical measures such as Neutrophil-to-Lymphocyte and Urea-to-Creatinine Ratios (Solimando et al. 2021) and IL-6 (Vultaggio et al. 2020) for the short-term prediction of COVID-19 patients status. Although these studies focused on the prediction of group level status changes, used CT for the prediction, or tracked the changes of medical measures, the individual level model predicting the ICU admission of a COVID-19 patient is not proposed.

Availability of Samples and Missing Data

For the prediction of patients' medical prognosis, studies have used medical records. However, in many cases, medical data has the missing values created by irregular observation of patient status and varying measurement frequency (Lipton et al. 2016). It is known that missing values are correlated with the target variable (Rubin 1976). Therefore, a variety of missing-value imputation techniques such as forward/backward-filling, zero-imputation, regression imputation, multiple imputation (Rubin 1996), Multivariate Imputation by Chained Equation (MICE; sBuuren and Groothuis-Oudshoorn 2010), interpolation (Kreindler and Lumsden 2012), spline methods (De Boor and De Boor 1978, p. 27), k-nearest neighbors (k-NN) algorithm (Song et al. 2008), and Expectation-Maximization (EM) algorithm (García-Laencina et al. 2010) have been widely applied in practice.

In the medical data domain, Lipton, Kale, and Wetzel (2016) proposed a new method by combining forward-filling and zero imputation to impute the missing values in electronic health records (EHRs). In their approach, binary indicators were used to represent that the value is imputed by an imputation strategy. They also indicated that recurrent neural networks (RNN) might recognize the fill-in values without indicators by learning the patterns of the fill-in values. Che et al. (2018) proposed GRU-D, which receives EHR-related variables, masking, and time intervals as input. The masking and time intervals enable the model to provide better prediction performance.

BRR was also used to impute the missing values. In the approach in Mostafa et al. (2020), authors used values of complete features as inputs to fit a BRR model and used the fitted model to predict the missing values. BRR is formalized as follows (Mostafa et al. 2020):

$$v \sim N(\mu, \alpha) \tag{1}$$

where:

$$\mu = \beta X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_n$$
$$\beta \sim N\ (0, \lambda^{-1} I_n)$$
$$\alpha \sim gamma\ (\alpha_1, \alpha_2)$$
$$\lambda \sim gamma\ (\lambda_1, \lambda_2)$$

Here, v is the target feature following Gaussian distribution, where variance is α and mean is μ = βX, X is the independent features, β is the regression parameter of which independent Gaussian priors are variance $\lambda^{-1} I_n$ and mean 0, n is the number of independent features, λ and α are regularizing parameter of gamma distribution, and α_1, α_2, λ_1, and λ_2 are hyper-parameters of the gamma prior distributions.

XGBoost

In a classification task, although the missing value imputation is an important pre-processing step to improve the classification performance, the selection of appropriate classification algorithm is critical. In recent years, XGBoost demonstrated successful performance in many competitions such as Kaggle (Ogunleye and Wang 2020) and attracted considerable attention from researchers and practitioners (Gumus and Kiran 2017). XGBoost is based on Gradient Boosting (Friedman 2001) and tree learners. It was proposed by Chen and Guestrin (2016) to improve the performance of Gradient Boosting. For output prediction, XGBoost uses *K* additive functions as follows (Chen and Guestrin 2016; Nobre and Neves 2019):

$$\hat{y}_i = \emptyset(x_\mathrm{i}) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{2}$$

where f is a function in the functional space $\mathcal{F}$, with $\mathcal{F} = \{f(x) = w_{q(x)}\}$ $(q: \mathbb{R}^m \rightarrow T,\ w \in \mathbb{R}^T)$ being the space of regression trees, where q is the structure of each tree that maps an example to the corresponding leaf index, T is the number of leaves in the tree, and w is the leaf weight.

Furthermore, it uses a loss function including a regularization term. The objective function $\mathcal{L}^{(t)}$ and its regularization term $\Omega(f_t)$ are formalized as follows (Chen and Guestrin 2016; Nobre and Neves 2019):

$$\mathcal{L}^{(t)} = \sum_{i}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{3}$$

In the objective function, $\hat{y}_i^{(t-1)}$ represents the prediction of the instance i at iteration t-1, $l(y_i, \hat{y}_i^{(t-1)})$ is the training loss function, and f_t is added to help the minimization of the objective (Nobre and Neves 2019). The second term Ω is for the regularization. It helps avoid overfitting by penalizing the complexity of the model (Chen and Guestrin 2016). Full details of XGBoost and its features are presented in Chen and Guestrin (2016).

XGBoost is adopted in various studies. In the business domain, XGBoost was used to predict the trade in the finance market (Nobre and Neves 2019) and crude oil price (Gumus and Kiran 2017). In the medical field, XGBoost was adopted for the classification of epilepsy patients (Torlay et al. 2017) and the diagnosis of Chronic Kidney Disease Diagnosis (Ogunleye and Wang 2020). Dhaliwal et al. (2018) used XGBoost for the Intrusion Detection System. These empirical results imply that XGBoost is a viable classifier for the various types of data created from diverse domains.

DATA DESCRIPTION

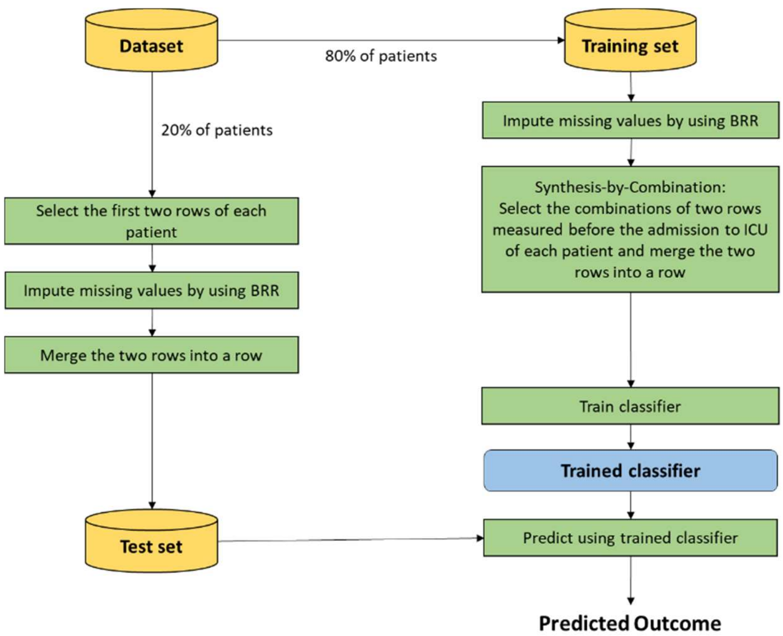
The dataset used in our experiment is a collection of anonymized COVID-19 inpatients' time series records of medical measurements from Hospital Sírio-Libanês, São Paulo and Brasilia. The data was anonymized by following international practices and posted at Kaggle (https://www.kaggle.com/S%C3%ADrio-Libanes/covid19).

The purpose of the creation of the dataset is to predict the likelihood of ICU admission of COVID-19 inpatients. The dataset is collected during five time windows. The time windows used are: from 0 to 2 hours, 2 to 4 hours, 4 to 6 hours, 6 to 12 hours, and above 12 hours from the hospital admission. As medical staff conduct blood tests and measure vital signs until the patient is admitted to ICU or released from the hospital, each patient's measurement records appear in five rows. However, 55.4% values in the dataset are missing because blood tests and vital sign checks are conducted irregularly depending on the patient's status. The number of COVID-19 patients in the dataset is 384. Each row has 54 features including patient demographic information (three features), patient previous grouped diseases (nine features), blood test results (36 features), and vital signs (six features) during a particular time window. These were expanded to 228 variables when pertinent to the mean, median, max, min, difference of max and min, and relative difference (= difference/median).

METHOD

The goal of the study is to propose a method predicting the likelihood of a COVID-19 inpatient's ICU admission based on the first two observations of the patient's status. The proposed method uses three steps to train the classifier. Three steps are the imputation of the missing values by applying BRR to each feature of each patient, the synthesis of training samples using the combinations of two rows measured before each patient is admitted to the ICU (simply, Synthesis-by-Combination), and XGBoost classifier. In the proposed method, any feature selection or variable importance selection method is not applied. The high-level view of our method is provided in Figure 1.

Figure 1: Method of the study



Imputation of Missing Values

As the first step of the proposed method, the missing values are imputed by fitting the BRR models based on Gaussian probability distribution (Massaoudi et al. 2020) because it is adaptable to insufficient data. Suppose there are missing values of the feature i among n features of a patient p, $x_{p,i}^{miss}$ and complete values of the feature of the patient, $x_{p,i}^{comp}$. The imputation of missing values of the feature i of patient p using BRR is formulated as follows:

$$x_{p,i}^{miss} \sim N(\mu, \alpha) \tag{4}$$

where:

$$\mu = \beta X = \beta_0 + \beta_i x_{p,i}^{comp}$$
$$\beta \sim N\,(0, \lambda^{-1} I_n)$$
$$\alpha \sim gamma\,(\alpha_1, \alpha_2)$$
$$\lambda \sim gamma\,(\lambda_1, \lambda_2)$$

Here, β is the regression parameter of which independent Gaussian priors are mean 0 and variance $\lambda^{-1} I_n$, n is the number of independent features n=1, λ and α are regularizing parameters of gamma distribution set to $\alpha_1 = \alpha_2 = \lambda_1 = \lambda_2 = 10^{-6}$. The entire imputation algorithm of the missing values is provided in Table 1.

Table 1: Algorithm for the imputation of missing values

Input:
  D: Dataset having complete and missing values
Output:
  $D_{comp+imp}$: Dataset with all missing values imputed

Definitions:
  P_{all}: List of all patients in the dataset
  D_p: Rows of patient *p*
  $\mathfrak{B}_{p,i}$: Bayesian Ridge Regression model for the feature *i* of patient *p*
  $x_{p,i}^{comp}$: Complete values in the feature *i* of patient *p*
  $x_{p,i}^{miss}$: Missing values in the feature *i* of patient *p*
  $x_{p,i}^{comp+imp}$: All data values of feature *i* of patient *p* in which all missing values are imputed

Begin
  foreach patient $p \in P_{all}$ **do**
    select D_p from *D*
    foreach feature *i* in D_p **do**
      fit $\mathfrak{B}_{p,i}$ using the values in $x_{p,i}^{comp}$
      $x_{p,i}^{comp+imp}$ ← **impute** $x_{p,i}^{miss}$ **using** $\mathfrak{B}_{p,i}$
      append $x_{p,i}^{comp+imp}$ **to** $D_{comp+imp}$
    end foreach
  end foreach
End

Synthesis of Training Samples Using Combination (Synthesis-by-Combination)

In the second step, we synthesize training samples by resampling the combinations of two rows from the rows created before each patient's ICU admission, maintaining the chronological order between two examples. When the two rows are concatenated, the patient ID and the time window features are removed from the two rows, so that the classifier could learn the samples without considering the patient ID and the time window information. As this combination method combines two rows of each patient, if only one row is created from a patient before ICU admission, the patient's records are not considered for the training. The algorithm used for the creation of combinations of rows to create a training set is in Table 2:

Table 2: Algorithm for the synthesis of training samples

Input:
  $D_{comp+imp}$: Dataset with all missing values imputed
Output:
  D_{synth}: Dataset synthesized using combination of two rows

Definitions:
  $P_{comp+imp}$: List of all patients in the dataset in which all missing values are imputed
  $D_{p,before}$: Rows of patient *p* created before the admission to ICU
  $D_{p,before}^{comb}$: All possible combinations of two rows created $D_{p,before}$

r_{first}: the row considered the first measurement of a patient
r_{second}: the row considered the second measurement of a patient

Begin
foreach patient $p \in P_{comp+\ imp}$ **do**
select $D_{p,\ before}$ from $D_{comp+\ imp}$
if the number of rows in $D_{p,\ before}$ >= 2 **do**
foreach combination of rows c in $D_{p,before}^{comb}$ **do**
$r_{first} \leftarrow$ earlier time window row in c
$r_{first} \leftarrow$ **remove** the patient ID and the time window features **from** r_{first}

$r_{second} \leftarrow$ later time window row in c
$r_{second} \leftarrow$ **remove** the patient ID and the time window features **from** r_{second}

append concatenate (r_{first}, r_{second}) **to** D_{synth}

end foreach
end if
end foreach
End

XGBoost Binary Classifier

For the final decision of the method, the XGBoost binary classifier is responsible for the classification. Among various classifiers, XGBoost is adopted considering the algorithm features. BRR and Synthesis-by-Combination steps in our method not only create the additional information for classifier training but also introduce noises. In addition, the number of patients available for the training is 260, which is a relatively small number of patients. Therefore, we consider the classifier that could improve the overall classification performance using multiple weak learners while preventing the overfitting. XGBoost uses weak learner sub-models in an additive way for the overall performance improvement and also uses the column sub-sampling and the regularization for the prevention of overfitting. As these features of XGBoost satisfied our criteria, we selected XGBoost as the classifier of the proposed method. The output of the classifier is the prediction of a COVID-19 patient's ICU admission and is within the range [0,1]. Scikit-learn machine learning library is used to implement the method. We did not conduct any hyper-parameter tuning and use the default hyper-parameter values of the library so that we could compare the performance of XGBoost with those of other classifiers.

EXPERIMENT AND DISCUSSION

Experimental Setup

In the experiment, the performance of our method is compared against the other methods combining three other imputation techniques, the Synthesis-by-Combination technique, and five other classifiers.

Three imputation techniques adopted for the experiment are the forward-filling and zero imputation (Lipton et al. 2016), the forward-filling and backward-filling, and the backward-filling and forward-filling. All imputation techniques employed are based on the same assumption that missing values would be the same to previous or subsequent values even if the strategies for filling the remaining missing values after the application of the first technique are different. The six classifiers selected to be compared with XGBoost are decision tree, SVM, Multi-layer

Perceptron (MLP), RNN, AdaBoost, and random forest. All methods combining the imputation techniques, the Synthesis-by-Combination technique, and classifiers in the experiment are listed in Table 3.

Among COVID-19 patients in the original dataset, we select the patients having at least two sets of measurements (two rows) measured before admission to the ICU. This is because the goal of our method is to predict the likelihood of ICU admission for each patient from the first two sets of the patient's measurements. The number of available patients is 325. From these 325 patients, we select 20% of the patients who have at least one measured value for each feature in two rows, in the order of more values in two rows, and use them as the test set. When creating the test set, we do not apply the Synthesis-by-Combination technique to be able to simulate the situation when a patient arrives at the hospital and has the status measured two times. The other 80% patients are used as the training set. The performance of the model is measured by using AUC-ROC. All classifiers, the imputation techniques, and the Synthesis-by-Combination technique in the experiment are implemented using Scikit-learn machine learning library and Python.

Experimental Results

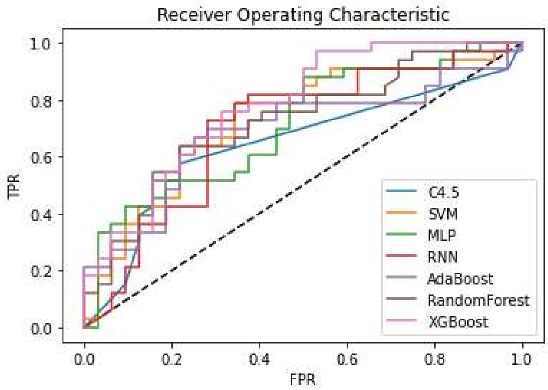
In the experiment, our method demonstrated better performance than all competing methods. Results are presented in Table 3.

Table 3: Results of Experiment

	Forward fill & zero imputation + Synthesis-by-Combination	Forward fill & backward fill imputation + Synthesis-by-Combination	Backward fill & forward fill imputation + Synthesis-by-Combination	BRR imputation + Synthesis-by-Combination
Decision Tree	0.646	0.716	0.584	0.544
SVM	0.729	0.780	0.750	0.738
MLP	0.709	0.725	0.783	0.739
RNN	0.701	0.700	0.669	0.629
AdaBoost	0.688	0.660	0.742	0.746
Random Forest	0.722	0.720	0.762	0.744
XGBoost	0.778	0.789	0.768	**0.823**

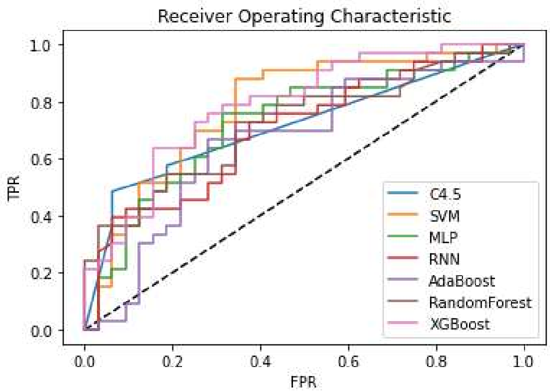
The first imputation technique we apply is the forward-filling and zero imputation used in Lipton, Kale, and Wetzel (2016). Among the methods using the forward-filling and zero imputation + Synthesis-by-Combination techniques, the method using XGBoost as the classifier outperforms other models (AUC-ROC=0.778). The AUC-ROCs of methods using the forward-filling and zero imputation + Synthesis-by-Combination techniques are depicted in Figure 2.

Figure 2: AUC-ROCs of the classifiers using forward-filling and zero imputation + Synthesis-by-Combination



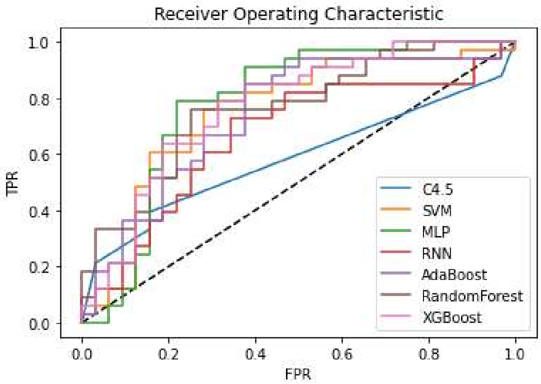
The second imputation method, forward-filling and backward-filling, improves the AUC-ROCs from forward-filling and zero imputation in four classifiers (decision tree, SVM, MLP, and XGBoost). Among seven classifiers combined with the forward-filling and backward-filling + Synthesis-by-Combination, XGBoost outperforms other methods (AUC-ROC=0. 789). The AUC-ROCs of methods using the forward-filling and backward-filling + Synthesis-by-Combination are depicted in Figure 3.

Figure 3: AUC-ROCs of the classifiers using forward-filling and backward-filling + Synthesis-by-Combination



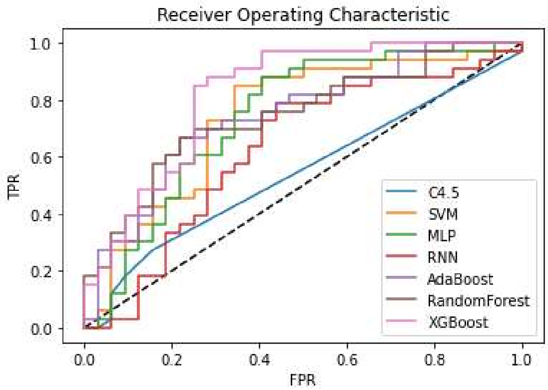
The third imputation technique, backward-filling and forward-filling, provides better performance than forward-filling and zero imputation in four classifiers (SVM, MLP, AdaBoost, and random forest). Among the results using the backward-filling and forward-filling imputation, MLP achieved the highest performance (AUC-ROC=0. 783). The AUC-ROCs of methods using the backward-filling and forward-filling + Synthesis-by-Combination are depicted in Figure 4.

Figure 4: AUC-ROCs of the classifiers using backward-filling and forward-filling + Synthesis-by-Combination



The imputation technique of the proposed method, BRR imputation, improves the AUC-ROCs of five classifiers except for the decision tree and RNN as compared to the methods using forward-filling and zero imputation technique. Among the methods using the BRR imputation, XGBoost achieves the highest performance (AUC-ROC=0. 823). The AUC-ROCs of methods using the BRR imputation + Synthesis-by-Combination are depicted in Figure 5.

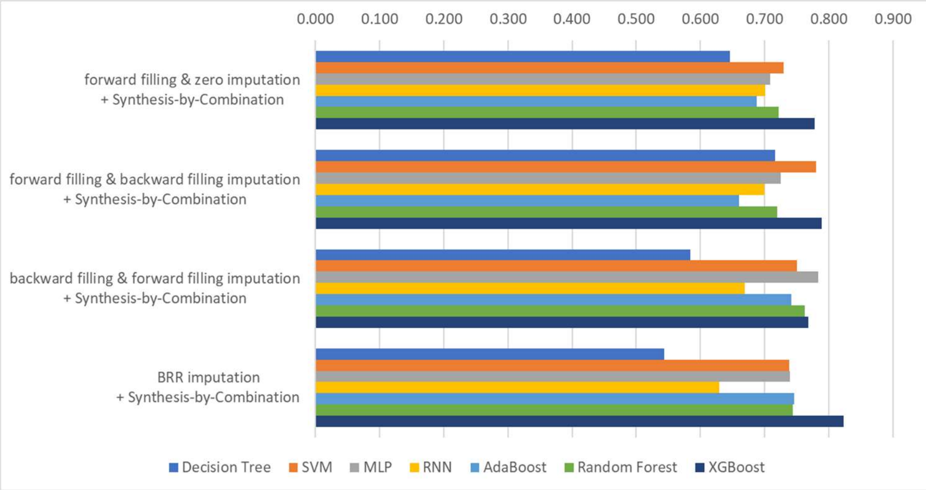
Figure 5: AUC-ROCS of the classifiers using BRR imputation + Synthesis-by-Combination



Discussion

In the experiment, we test a method combining three steps: BRR imputation, Synthesis-by-Combination, and XGBoost classifier. The AUC-ROC of the proposed method is compared with those of all other methods and the proposed method outperforms other methods (AUC-ROC=0.823). While many methods provide an AUC-ROC higher than 0.700, only our method provides an AUC-ROC higher than 0.8. The AUC-ROCs of all methods in the experiment are depicted in Figure 6.

Figure 6. AUC-ROCs of models in the experiment



Among the results of methods combining BRR imputation + Synthesis-by-Combination with non-XGBoost classifiers and the results of methods combining the non-BRR imputation + Synthesis-by-Combination with XGBoost classifier, any methods did not achieve the AUC-ROC of our method (AUC-ROC = 0.823). That implies that the proposed method, BRR imputation + Synthesis-by-Combination + XGBoost, is the best combination to achieve the highest performance of this study while other methods also improved AUC-ROC.

CONCLUSION

During the COVID-19 outbreak, hospitals suffered from the surge of patients requiring ICU admission. Aziz et al. (2020) suggested using mathematical modeling for improved organizational management. In this study, we develop a method predicting the individual COVID-19 inpatient's likelihood of ICU admission within several hours based on two sets of patient measurements. We use the dataset collected during the early stage of the outbreak of COVID-19 from a hospital in Brazil.In the dataset, 325 patients are available and 55.4% of values are missing, so a method with two pre-processing steps is proposed. In the first step, the missing values are imputed by an imputation technique using BRR. In the second step, we synthesize new samples by using the combinations of two rows of each patient to improve the variability of the training samples. As the final step of the method, XGBoost classifier is adopted to predict the likelihood of ICU admission for each patient.

In the experiment, we compare the AUC-ROC of our method to those of others using various imputation techniques and classifiers (but the same training sample synthesizing method). Among the various methods in the experiment, our method provides the highest AUC-ROC (AUC-ROC = 0.823). The comparison of the results from various combinations of imputation techniques and classifiers demonstrates that the proposed method of this study is the best combination to achieve the performance of our method.

Although our method demonstrated good performance, there is room for improvements in the future. There exist many missing values in medical data owing to the characteristics of medical record creation procedures. The imputation techniques for medical data considering the characteristics of medical procedure should be developed or further improved because better imputation techniques have potential to improve the model performance using medical data. In addition, if we could obtain more detailed data from the hospital and apply a feature selection method, it would be helpful to provide more understandable and actionable decision support to the medical staff.

REFERENCES

Assaf, D., Gutman, Y., Neuman, Y., Segal, G., Amit, S., Gefen-Halevi, S., Shilo, N., Epstein, A., Mor-Cohen, R., Biber, A., Rahav, G., Levy, I., and Tirosh, A. 2020. "Utilization of Machine-Learning Models to Accurately Predict the Risk for Critical COVID-19," *Internal and Emergency Medicine* (15:8), pp. 1435–1443. (https://doi.org/10.1007/s11739-020-02475-0).

Aziz, S., Arabi, Y. M., Alhazzani, W., Evans, L., Citerio, G., Fischkoff, K., Salluh, J., Meyfroidt, G., Alshamsi, F., Oczkowski, S., Azoulay, E., Price, A., Burry, L., Dzierba, A., Benintende, A., Morgan, J., Grasselli, G., Rhodes, A., Møller, M. H., Chu, L., Schwedhelm, S., Lowe, J. J., Bin, D., and Christian, M. D. 2020. "Managing ICU Surge during the COVID-19 Crisis: Rapid Guidelines," *Intensive Care Medicine* (46:7), pp. 1303–1325. (https://doi.org/10.1007/s00134-020-06092-5).

Bekker, R., and Koole, G. 2021. "Modeling COVID-19 Hospital Admissions and Occupancy in the Netherlands," *ArXiv Preprint ArXiv:2102.11021*.

Berta, P., Paruolo, P., Verzillo, S., and Lovaglio, P. G. 2020. "A Bivariate Prediction Approach for Adapting the Health Care System Response to the Spread of COVID-19," *PLOS ONE* (15:10), (T. Mathes, ed.), p. e0240150. (https://doi.org/10.1371/journal.pone.0240150).

Bonnasse-Gahot, L., Dénès, M., Dulac-Arnold, G., Girgin, S., Husson, F., Iovene, V., Josse, J., Kimmoun, A., Landes, F., and Nadal, J.-P. 2020. *ICU Bed Availability Monitoring and Analysis in the Grand Est Region of France during the COVID-19 Epidemic*.

Burdick, H., Lam, C., Mataraso, S., Siefkas, A., Braden, G., Dellinger, R. P., McCoy, A., Vincent, J.-L., Green-Saxena, A., Barnes, G., Hoffman, J., Calvert, J., Pellegrini, E., and Das, R. 2020. "Prediction of Respiratory Decompensation in Covid-19 Patients Using Machine Learning: The READY Trial," *Computers in Biology and Medicine* (124), p. 103949. (https://doi.org/10.1016/j.compbiomed.2020.103949).

Buuren, S. van, and Groothuis-Oudshoorn, K. 2010. "Mice: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software*, University of California, Los Angeles, pp. 1–68.

Català, M., Alonso, S., Alvarez-Lacalle, E., López, D., Cardona, P.-J., and Prats, C. 2020. "Empirical Model for Short-Time Prediction of COVID-19 Spreading," *PLOS Computational Biology* (16:12), (B. M. Althouse, ed.), p. e1008431. (https://doi.org/10.1371/journal.pcbi.1008431).

Chassagnon, G., Vakalopoulou, M., Battistella, E., Christodoulidis, S., Hoang-Thi, T.-N., Dangeard, S., Deutsch, E., Andre, F., Guillo, E., and Halm, N. 2020. "AI-Driven CT-Based Quantification, Staging and Short-Term Outcome Prediction of COVID-19 Pneumonia," *ArXiv Preprint ArXiv:2004.12852*.

Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. 2018. "Recurrent Neural Networks for Multivariate Time Series with Missing Values," *Scientific Reports* (8:1), p. 6085. (https://doi.org/10.1038/s41598-018-24271-9).

Chen, T., and Guestrin, C. 2016. "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. (https://doi.org/10.1145/2939672.2939785).

Chin, V., Samia, N. I., Marchant, R., Rosen, O., Ioannidis, J. P., Tanner, M. A., and Cripps, S. 2020. "A Case Study in Model Failure? COVID-19 Daily Deaths and ICU Bed Utilisation Predictions in New York State," *European Journal of Epidemiology* (35:8), Springer, pp. 733–742.

De Boor, C., and De Boor, C. 1978. *A Practical Guide to Splines*, (Vol. 27), springer-verlag New York.

Dhaliwal, S., Nahid, A.-A., and Abbas, R. 2018. "Effective Intrusion Detection System Using XGBoost," *Information* (9:7), p. 149. (https://doi.org/10.3390/info9070149).

El-Ghitany, E. 2020. "A Short-Term Forecast Scenario for COVID-19 Epidemic and Allocated Hospital Readiness in Egypt," preprint, Preprint, MEDICINE & PHARMACOLOGY, April 27. (https://doi.org/10.20944/preprints202004.0473.v1).

Farcomeni, A., Maruotti, A., Divino, F., Jona-Lasinio, G., and Lovison, G. 2021. "An Ensemble Approach to Short-term Forecast of COVID-19 Intensive Care Occupancy in Italian Regions," *Biometrical Journal* (63:3), pp. 503–513. (https://doi.org/10.1002/bimj.202000189).

Francone, M., Iafrate, F., Masci, G. M., Coco, S., Cilia, F., Manganaro, L., Panebianco, V., Andreoli, C., Colaiacomo, M. C., Zingaropoli, M. A., Ciardi, M. R., Mastroianni, C. M., Pugliese, F., Alessandri, F., Turriziani, O., Ricci, P., and Catalano, C. 2020. "Chest CT Score in COVID-19 Patients: Correlation with Disease Severity and Short-Term Prognosis," *European Radiology* (30:12), pp. 6808–6817. (https://doi.org/10.1007/s00330-020-07033-y).

Friedman, J. H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine.," *The Annals of Statistics* (29:5). (https://doi.org/10.1214/aos/1013203451).

Funk, S., Abbott, S., Atkins, B., Baguelin, M., Baillie, J., Birrell, P., Blake, J., Bosse, N., Burton, J., Carruthers, J., Davies, N., De Angelis, D., Dyson, L., Edmunds, W., Eggo, R., Ferguson, N., Gaythorpe, K., Gorsich, E., Guyver-Fletcher, G., Hellewell, J., Hill, E., Holmes, A., House, T., Jewell, C., Jit, M., Jombart, T., Joshi, I., Keeling, M., Kendall, E., Knock, E., Kucharski, A., Lythgoe, K., Meakin, S., Munday, J., Openshaw, P., Overton, C., Pagani, F., Pearson, J., Perez-Guzman, P., Pellis, L., Scarabel, F., Semple, M., Sherratt, K., Tang, M., Tildesley, M., Van Leeuwen, E., Whittles, L., CMMID COVID-19 Working Group, Imperial College COVID-19 Response Team, and ISARIC4C Investigators. 2020. "Short-Term Forecasts to Inform the Response to the Covid-19 Epidemic in the UK," preprint, Preprint, Epidemiology, November 13. (https://doi.org/10.1101/2020.11.11.20220962).

Gao, Y., Cai, G.-Y., Fang, W., Li, H.-Y., Wang, S.-Y., Chen, L., Yu, Y., Liu, D., Xu, S., Cui, P.-F., Zeng, S.-Q., Feng, X.-X., Yu, R.-D., Wang, Y., Yuan, Y., Jiao, X.-F., Chi, J.-H., Liu, J.-H., Li, R.-Y., Zheng, X., Song, C.-Y., Jin, N., Gong, W.-J., Liu, X.-Y., Huang, L., Tian, X., Li, L., Xing, H., Ma, D., Li, C.-R., Ye, F., and Gao, Q.-L. 2020. "Machine Learning Based Early Warning System Enables Accurate Mortality Risk Prediction for COVID-19," *Nature Communications* (11:1), p. 5033. (https://doi.org/10.1038/s41467-020-18684-2).

García-Laencina, P. J., Sancho-Gómez, J.-L., and Figueiras-Vidal, A. R. 2010. "Pattern Classification with Missing Data: A Review," *Neural Computing and Applications* (19:2), Springer, pp. 263–282.

Goic, M., Bozanic-Leal, M. S., Badal, M., and Basso, L. J. 2021. "COVID-19: Short-Term Forecast of ICU Beds in Times of Crisis," *PLOS ONE* (16:1), (C.-H. Chen, ed.), p. e0245272. (https://doi.org/10.1371/journal.pone.0245272).

Gomes, D. P. S., Ulhaq, A., Paul, M., Horry, M. J., Chakraborty, S., Saha, M., Debnath, T., and Rahaman, D. M. M. 2020. "Potential Features of ICU Admission in X-Ray Images of COVID-19 Patients," *ArXiv:2009.12597 [Cs, Eess]*. (http://arxiv.org/abs/2009.12597).

Gumus, M., and Kiran, M. S. 2017. "Crude Oil Price Forecasting Using XGBoost," in *2017 International Conference on Computer Science and Engineering (UBMK)*, IEEE, pp. 1100–1103.

He, F., Quan, Y., Lei, M., Liu, R., Qin, S., Zeng, J., Zhao, Z., Yu, N., Yang, L., and Cao, J. 2020. "Clinical Features and Risk Factors for ICU Admission in COVID-19 Patients with Cardiovascular Diseases," *Aging and Disease* (11:4), p. 763. (https://doi.org/10.14336/AD.2020.0622).

Keeling, M. J., Dyson, L., Guyver-Fletcher, G., Holmes, A., Semple, M. G., ISARIC4C Investigators, Tildesley, M. J., and Hill, E. M. 2020. "Fitting to the UK COVID-19 Outbreak, Short-Term Forecasts and Estimating the Reproductive Number," preprint, Preprint, Infectious Diseases (except HIV/AIDS), August 4. (https://doi.org/10.1101/2020.08.04.20163782).

Kreindler, D. M., and Lumsden, C. J. 2012. "The Effects of the Irregular Sample and Missing Data in Time Series Analysis," *Nonlinear Dynamical Systems Analysis for the Behavioral Sciences Using Real Data* (135:2).

Lipton, Z. C., Kale, D. C., and Wetzel, R. 2016. "Modeling Missing Data in Clinical Time Series with RNNs," *ArXiv:1606.04130 [Cs, Stat]*. (http://arxiv.org/abs/1606.04130).

Manevski, D., Ružić Gorenjec, N., Kejžar, N., and Blagus, R. 2020. "Modeling COVID-19 Pandemic Using Bayesian Analysis with Application to Slovene Data," *Mathematical Biosciences* (329), p. 108466. (https://doi.org/10.1016/j.mbs.2020.108466).

Massaoudi, M., Refaat, S. S., Abu-Rub, H., Chihi, I., and Wesleti, F. S. 2020. "A Hybrid Bayesian Ridge Regression-CWT-Catboost Model For PV Power Forecasting," in *2020 IEEE Kansas Power and Energy Conference (KPEC)*, Manhattan, KS, USA: IEEE, July, pp. 1–5. (https://doi.org/10.1109/KPEC47870.2020.9167596).

Massonnaud, C., Roux, J., and Crépey, P. 2020. “COVID-19: Forecasting Short Term Hospital Needs in France,” preprint, Preprint, Epidemiology, March 20. (https://doi.org/10.1101/2020.03.16.20036939).

Mostafa, S. M., Eladimy, A. S., Hamad, S., and Amano, H. 2020. “CBRG: A Novel Algorithm for Handling Missing Data Using Bayesian Ridge Regression and Feature Selection Based on Gain Ratio,” *IEEE Access* (8), pp. 216969–216985. (https://doi.org/10.1109/ACCESS.2020.3042119).

Nobre, J., and Neves, R. F. 2019. “Combining Principal Component Analysis, Discrete Wavelet Transform and XGBoost to Trade in the Financial Markets,” *Expert Systems with Applications* (125), pp. 181–194. (https://doi.org/10.1016/j.eswa.2019.01.083).

Ogunleye, A., and Wang, Q.-G. 2020. “XGBoost Model for Chronic Kidney Disease Diagnosis,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (17:6), pp. 2131–2140. (https://doi.org/10.1109/TCBB.2019.2911071).

Petermann, M., Frhlich, J., and Wyler, D. 2021. “A Case Study of Short-Term Forecasting during the COVID-19 Pandemic: The Platform Icumonitoring.Ch,” *Swiss Medical Weekly*. (https://doi.org/10.4414/smw.2021.20524).

Rahaman Khan, M. H., and Hossain, A. 2020. “COVID-19 Outbreak Situations in Bangladesh: An Empirical Analysis,” preprint, Preprint, Infectious Diseases (except HIV/AIDS), April 22. (https://doi.org/10.1101/2020.04.16.20068312).

Ricoca Peixoto, V., Vieira, A., Aguiar, P., Carvalho, C., Thomas, D., and Abrantes, A. 2020. “Rapid Assessment of the Impact of Lockdown on the COVID-19 Epidemic in Portugal,” preprint, Preprint, Epidemiology, May 27. (https://doi.org/10.1101/2020.05.26.20098244).

Roncon, L., Zuin, M., Rigatelli, G., and Zuliani, G. 2020. “Diabetic Patients with COVID-19 Infection Are at Higher Risk of ICU Admission and Poor Short-Term Outcome,” *Journal of Clinical Virology* (127), p. 104354. (https://doi.org/10.1016/j.jcv.2020.104354).

Rubin, D. B. 1976. “Inference and Missing Data,” *Biometrika* (63:3), Oxford University Press, pp. 581–592.

Rubin, D. B. 1996. “Multiple Imputation after 18+ Years,” *Journal of the American Statistical Association* (91:434), Taylor & Francis Group, pp. 473–489.

Solimando, A. G., Susca, N., Borrelli, P., Prete, M., Lauletta, G., Pappagallo, F., Buono, R., Inglese, G., Forina, B. M., Bochicchio, D., Capobianco, M., Carrieri, V., Cicco, S., Leone, P., Silvestris, N., Saracino, A., Ria, R., Procacci, V., Migliore, G., Vacca, A., and Racanelli, V. 2021. “Short-Term Variations in Neutrophil-to-Lymphocyte and Urea-to-Creatinine Ratios Anticipate Intensive Care Unit Admission of COVID-19 Patients in the Emergency Department,” *Frontiers in Medicine* (7), p. 625176. (https://doi.org/10.3389/fmed.2020.625176).

Song, Q., Shepperd, M., Chen, X., and Liu, J. 2008. “Can K-NN Imputation Improve the Performance of C4. 5 with Small Software Project Data Sets? A Comparative Evaluation,” *Journal of Systems and Software* (81:12), Elsevier, pp. 2361–2370.

Sun, Liping, Song, F., Shi, N., Liu, F., Li, S., Li, P., Zhang, W., Jiang, X., Zhang, Y., Sun, Lining, Chen, X., and Shi, Y. 2020. “Combination of Four Clinical Indicators Predicts the Severe/Critical Symptom of Patients Infected COVID-19,” *Journal of Clinical Virology* (128), p. 104431. (https://doi.org/10.1016/j.jcv.2020.104431).

Torlay, L., Perrone-Bertolotti, M., Thomas, E., and Baciu, M. 2017. “Machine Learning–XGBoost Analysis of Language Networks to Classify Patients with Epilepsy,” *Brain Informatics* (4:3), pp. 159–169. (https://doi.org/10.1007/s40708-017-0065-7).

Vultaggio, A., Vivarelli, E., Virgili, G., Lucenteforte, E., Bartoloni, A., Nozzoli, C., Morettini, A., Berni, A., Malandrino, D., Rossi, O., Nencini, F., Pieralli, F., Peris, A., Lagi, F., Scocchera, G., Spinicci, M., Trotta, M., Mazzetti, M., Parronchi, P., Cosmi, L., Liotta, F., Fontanari, P., Mazzoni, A., Salvati, L., Maggi, E., Annunziato, F., Almerigogna, F., and Matucci, A. 2020. “Prompt Predicting of Early Clinical Deterioration of Moderate-to-Severe COVID-19 Patients: Usefulness of a Combined Score Using IL-6 in a Preliminary Study,” *The Journal of Allergy and Clinical Immunology: In Practice* (8:8), pp. 2575-2581.e2. (https://doi.org/10.1016/j.jaip.2020.06.013).

Weissman, G. E., Crane-Droesch, A., Chivers, C., Luong, T., Hanish, A., Levy, M. Z., Lubken, J., Becker, M., Draugelis, M. E., Anesi, G. L., Brennan, P. J., Christie, J. D., Hanson, C. W., Mikkelsen, M. E., and Halpern, S. D. 2020. “Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic,” *Annals of Internal Medicine* (173:1), pp. 21–28. (https://doi.org/10.7326/M20-1260).

Zhao, C., Tepekule, B., Criscuolo, N. G., Wendel Garcia, P. D., Hilty, M. P., RISC-19-ICU consortium Investigators in Switzerland, Fumeaux, T., and Van Boeckel, T. 2020. “Icumonitoring.Ch: A Platform for Short-Term Forecasting of Intensive Care Unit Occupancy during the COVID-19 Epidemic in Switzerland,” *Swiss Medical Weekly*. (https://doi.org/10.4414/smw.2020.20277).

Zhao, H., Merchant, N. N., McNulty, A., Radcliff, T. A., Cote, M. J., Fischer, R. S. B., Sang, H., and Ory, M. G. 2021. “COVID-19: Short Term Prediction Model Using Daily Incidence Data,” *PLOS ONE* (16:4), (J. Schieffelin, ed.), p. e0250110. (https://doi.org/10.1371/journal.pone.0250110).

Zhao, Z., Chen, A., Hou, W., Graham, J. M., Li, H., Richman, P. S., Thode, H. C., Singer, A. J., and Duong, T. Q. 2020. “Prediction Model and Risk Scores of ICU Admission and Mortality in COVID-19,” *PLOS ONE* (15:7), (M. Adrish, ed.), p. e0236618. (https://doi.org/10.1371/journal.pone.0236618).

DECISION SCIENCES INSTITUTE

A Case Study of Effectiveness Analysis and Simulation of the COVID-19 Prevention Policies Based on Population Dynamics Modeling

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Yiyun (Cindy) Fei
Georgia Institute of Technology
Email: yfei38@gatech.edu

Jianyuan Peng
Georgia Institute of Technology
Email: pjy@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

The paper proposed a population dynamics model to simulate the COVID-19 pandemic and analyze the effectiveness of prevention policies in the early stage. The model is designed to aid the decision-making process of policy-making in the early stage. The model is formulated based on the SEIR model to simulate the spread of COVID-19 from human to human. By implementing the data in the U.S., the model is first fitted to the data first. Then, the model simulates the number of infected people with the change of time under different levels of social distancing and mask-wearing.

KEYWORDS: Population Dynamics, COVID-19, Model-Fitting, Simulation

INTRODUCTION

In December 2019, several cases of coronavirus virus symptoms were observed in Wuhan City, Hubei Province of China. On December 29th 2019, WHO's Country Office in China reported the case of novel coronavirus pneumonia (World Health Organization, 2020). On February 11st 2019, the novel coronavirus was named coronavirus disease 2019 (COVID-19) officially by WHO (World Health Organization, 2020). The virus that causes COVID-19 is named "severe acute respiratory syndrome coronavirus 2" (SARS-Cov-2) by the Coronavirus Study Group, a single-standard RNA beta-coronavirus (Lotfi et al., 2020). The COVID-19 spreads fast worldwide. As of May 26th 2021, around 168 million confirmed cases of COVID-19, with around 3.5 million death cases. There are around 66 million confirmed infected cases in the American area, with around 20 thousand weekly increases by May 2021(WHO, 2021).

The SARS-CoV-2 transmits through human-to-human, by respiratory droplets, and contaminated objects. Respiratory droplets is the main human-to-human spread approach since it can crisscross less than 6 feet and stay in the area at a short period (Lotfi et al., 2020). The spread of COVID-19 draws researchers' and the government's attention. At the early stage, governments implemented strategies like lockdown, encouraging social distancing, recommending to wear masks to control and prevent the spread of COVID-19. Reducing the contact rate of people is the core of these strategies. Controls like isolations are done for the detected cases. In 2021, the successful development of the vaccine makes immunization for most people possible. As of May 2021, there are around 1.81 billion doses of the COVID-19 vaccine being administered (Mathieu et al., 2021). In February 2020, the first COVID-19 case is confirmed in the United States in the NewYork City. The COVID-19 spread rapidly in the United States. The confirmed COVID-19 cases rose to 395,926 by April 7th 2020 with 12,757 death cases and 21,602 daily new cases (Bialek et al., 2020). The daily increases of confirmed cases reached the peak on January 8th 2021 with the daily new cases of 303,549 new confirmed cases. The daily new cases decrease after the peak as the result of Vaccination. As of May 28th 2021, the daily new cases decrease to 22,138 (CDC, 2021).

Before the boosting level of immunity with using the vaccine, prevention and control of the COVID-19 pandemic depend on the prevention policies made by the government. However, prevention policies like lockdown, restrictions on business and activities, and banned gatherings have a huge effect on the economy (Zaremba et al., 2020). In the prevention policies and health decision-making process, the mathematical pandemic model is extremely significant (Egger et al., 2017). This paper presents a mathematical population dynamic model for calculating and predicting the number of infected cases with limited known parameters. The model is designed to aid the policy-making process in the early stage of the pandemic to control the spread of diseases and their effects on the economy. The rest of the paper includes the literature review of the population dynamic-based pandemic model, the model formulation and analysis, and a study case using the COVID-19 data of the U.S.

LITERATURE REVIEW

The SEIR model is widely used for simulating the transmission of diseases that spread from population to population through direct contact. In the most SEIR model, there are four classes: (1) The class $\boldsymbol{S}$ represents the susceptible population, (2) The class $\boldsymbol{E}$ represents the population that exposed to the infectious virus, (3) The class $\boldsymbol{I}$ represents the infected population, (4) The class $\boldsymbol{R}$ represents the recovered population (Li et al., 1999). Different type of SEIR model has been conducted by researchers to simulate the outcome of COVID-19.

A new mathematical model for COVID-19 population dynamics is proposed to analyze the effect of non-pharmaceutical control policies. By implementing the data in Lagos of Nigeria, the model can forecast the cumulative detected COVID-19 cases at different levels of control policies like social distancing or wearing masks (Okuonghae & Omame, 2020). Compared to the traditional SEIR model, the new model includes several new classes for better simulating the situation.

Another mathematical model is presented for analyzing the severity and potential of the COVID-19 pandemic. The presented model is modified from the basic SEIR model, where there is a proportion of some classes are quarantined and isolated. By implementing the data in Wuhan, China, The analysis results of the model stress the significance of the isolation and quarantine policies (Tang et al., 2020).

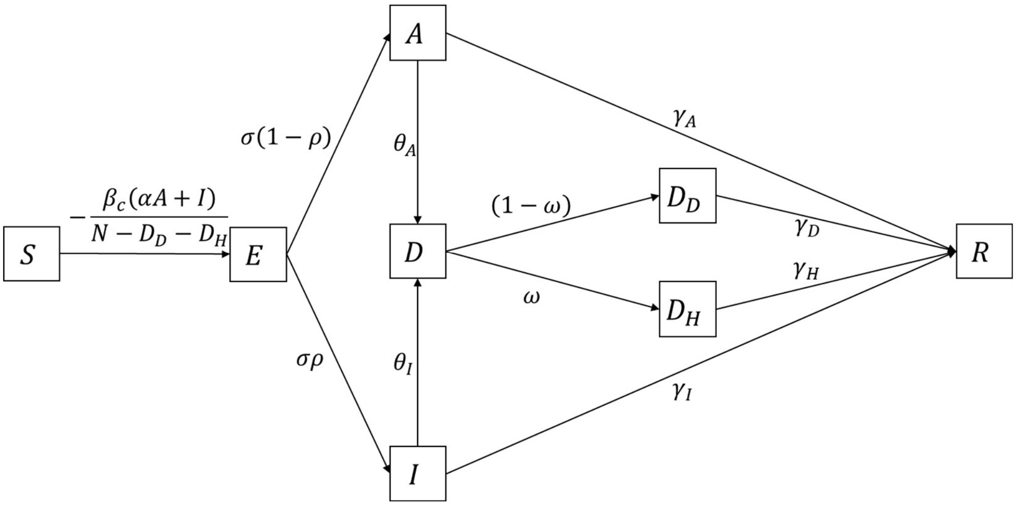
Despite researching the transmission network model between humans, a Bats-Hosts-Reservoir-People transmission network model is developed to simulate the possible spread process from bat to human (Chen et al., 2020). There are four sub-network in the proposed system which are Bats(the infection source), unknown hosts, reservoir, and people.

MODEL FORMULATION

The proposed model is an improved-SEIR model. the proposed model. It is assumed that birth and natural death are not included in the model, which means there aren't any demographic parameters. Figure 1 shows the network architecture of the proposed improved SEIR model. There are eight variables in the model which represents the number of eight types of groups of people with variable time $\boldsymbol{t}$: (1) $\boldsymbol{S}$ represents susceptible people, (2) $\boldsymbol{E}$ represents the number of exposed people, (3) $\boldsymbol{A}$ represents the number of undetected asymptomatically infected people, (4) $\boldsymbol{I}$ represents the number of undetected symptomatically infected people, (5) $\boldsymbol{D}$ represents the number of detected infected people, (6) $\boldsymbol{D_D}$ represents the number of home isolated infected people, (7) $\boldsymbol{D_H}$ represents the number of hospitalized isolated infected people and (8) $\boldsymbol{R}$ represents the number of recovered people. The total number of people in the system is $\boldsymbol{N(t)}$ where can be expressed as equation 1.

$$N(t) = S(t) + E(t) + A(t) + I(t) + D(t) + D_D(t) + D_H(t) + R(t) \quad (1)$$

Figure 1: The network diagram of the proposed model



The COVID-19 spread flow in the model starts from the susceptible people, and then the susceptible people transmit to the exposed people. Once a person in the exposed people gets infected, there is a chance that the person becomes asymptomatic or symptomatic. Asymptomatic and symptomatic people have different rates of being detected and different death rates. The detected infected people have the probability of being home isolated or hospital isolated. Asymptomatic people, symptomatic people, home isolated infected people, and hospitalized isolated infected people have the rates of being recovered. In this model, once a person is recovered, it can't get infected again.

Table 1: Description of the parameter in the model

Parameter	Description
β_c	Effective Contact Rate between S and E.
α	Parameter controls the infectious level of asymptomatic people.
σ	Transsition rate from exposed state to infectious people A and I.
ρ	Probability of Become Symptomatic Infectious
θ_A,θ_I	Detection rate of the infectious people A and I.
ω	Hospitalization rate of detected infected people.
$\gamma_A, \gamma_I, \gamma_D, \gamma_H$	Recover rate of the people A, D_D, D_H, and I
d_A, d_I, d_D, d_H	Death rate of the people A, D_D, D_H, and I

Table 1 shows the parameter that constructs different transmission rates. These parameters lie inside the interval of $[\mathbf{0}, \mathbf{1}]$. The transmission dynamics of the model is expressed in a system of non-linear differential equation as shown from equation 2 to equation 9. The system of differential equations describes the basic version of the proposed model.

$$S' = -\frac{\beta_c(\alpha A + I)}{N - I_D - I_H} \tag{2}$$

$$E' = \frac{\beta c(\alpha A + I)}{N - I_D - I_H} S - \sigma E \tag{3}$$

$$A' = \sigma(1 - \rho)E - (\theta_A + \gamma_A + d_A)A \tag{4}$$

$$I' = \sigma(\rho)E - (\theta_I + \gamma_I + d_I)I \tag{5}$$

$$D' = \theta_A A + \theta_I I - \lambda D \tag{6}$$

$$D_D' = (1 - \omega)D - \gamma_D I D_D - d_D D_D \tag{7}$$

$$D_H' = (\omega)D - \gamma_H D_H - d_h D_h \tag{8}$$

$$R' = \gamma_I I + \gamma_H D_H + \gamma_D D_D + \gamma_A A \tag{9}$$

The model could be modified for different using purposes by introducing more parameters. In this paper, two parameters are introduced and discussed, which are the parameter that represents social distancing $\boldsymbol{\nu}$ and the parameter that represents the level of wearing mask $\boldsymbol{\tau}$. These two parameters affected the transmission rate from susceptible people $\boldsymbol{S}$ to exposed people $\boldsymbol{E}$. Like other parameters in the model, these two parameters are between 0 to 1. Equation 2 and equation 3 in the basic model are also modified, as shown in equation 10 to equation 12.

$$S' = -\frac{\beta_{c\,mod}(\alpha A + I)}{N - I_D - I_H} \tag{10}$$

$$E' = \frac{\beta_{c\,mod}(\alpha A + I)}{N - I_D - I_H} S - \sigma E \tag{11}$$

$$\beta_{c\,mod} = \nu\tau\beta_c \tag{12}$$

MODEL ANALYSIS

Diseases free equilibrium (DFE) is significant for an epidemiological mathematical model to determine when the diseases become absent in the population (Pellis et al., 2012). In the proposed model, the basic reproduction number $\boldsymbol{\mathcal{R}_0}$ represents the expected value of the secondary case by typical infected individuals in a completed susceptible population (Diekmann et al., 1990). When the basic reproduction number $\boldsymbol{\mathcal{R}_0}$ is bigger than 1, the diseases can continually invade the population, otherwise, the DFE is stable. In this section, the basic reproduction number $\boldsymbol{\mathcal{R}_0}$ of the modified scenario that includes social distancing parameter $\boldsymbol{\nu}$ and the level of mask-wearing parameter $\boldsymbol{\tau}$ is introduced here since these two variables are assumed as the two responses policies in the study case.

$$F=\begin{bmatrix} 0 & \beta_c\alpha\,\nu\,\tau & \beta_c\nu\,\tau & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{13}$$

$$V=\begin{bmatrix} \omega & 0 & 0 & 0 & 0 & 0 \\ \omega\,(1-\rho\,) & (\theta_A+\gamma_A) & 0 & 0 & 0 & 0 \\ \omega\,\rho & 0 & \theta_I+\gamma_I+d_I & 0 & 0 & 0 \\ 0 & -\theta_A & -\theta_I & \lambda+d_A & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma_D+d_D & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma_H+d_h \end{bmatrix} \tag{14}$$

$$\mathcal{R}_0 = \rho(FV^{-1}) = \frac{\alpha\beta_c\nu\,\tau(\,\rho\,-\,1)}{\gamma_A\,+\theta_A} - \frac{\alpha\beta_c\nu\,\tau\rho}{d_I+\gamma_I+\theta_I} \tag{15}$$

By using the notation presented by Driessche et al. in 2002, the matrix $\boldsymbol{F}$ in equation 13 represents new infections in the model, and the matrix $\boldsymbol{V}$ in equation 14 represents the transmission between population groups. The expression of $\boldsymbol{\mathcal{R}_0}$ is shown in equation 15.

CASE STUDY: THE UNITED STATES CASE

In this section, the proposed model is employed for simulating the COVID-19 dynamics in the United States. Using the data of confirmed cases and recovered cases of the United States starts from January 22nd 2020, the model is fitted based on the given data. The time interval of the simulation is 325 days which stop before implementing the vaccines. The data is provided by an interactive web-based dashboard for tracking the real-time global COVID-19 cases (Dong et al., 2020). The social distancing parameter $\boldsymbol{\nu}$ and mask-wearing parameter $\boldsymbol{\tau}$ are employed for variable studies after the curve-fitting process using the basic proposed model.

Table 2: Parameter settings of the case study

Parameter	Value	Data Source
Parameter controls the infectious level of asymptomatic people: $\boldsymbol{\alpha}$	0.5	(Chen et al., 2020)
The effective contact rate: $\boldsymbol{\beta_c}$	Fitted	-
Probability of becoming symptomatic infectious: $\boldsymbol{\rho}$	42.50%	(Oran et al., 2020)
Transsition rate from exposed state to infectious State: $\boldsymbol{\sigma}$	$\frac{1}{5.1}$	(Lauer et al., 2020)
Hospitalization Rate : $\boldsymbol{\omega}$	0.2877	(CDC, 2021) (U.S. Census Bureau, 2019)
Death Rate of Hosptialized people: $\boldsymbol{d_D}$	0.01928	(CDC, 2020)
Death Rate of Outpatient: $\boldsymbol{d_A}, \boldsymbol{d_I}, \boldsymbol{d_{I_D}}$	0.015	(KRR et al., 2020)
Recover rate of the asympotmatic infected people: $\boldsymbol{\gamma_A}$	0.13978	(Tang et al., 2020)
Recover rate of the infected sypotamitc people: $\boldsymbol{\gamma_I}$	0.33029	(Tang et al., 2020)
Recover rate of the people D_D and D_H: $\boldsymbol{\gamma_H}, \boldsymbol{\gamma_D}$	Fitted	-
Detection rate: $\boldsymbol{\theta_A}, \boldsymbol{\theta_I}$	Fitted	-

The value of parameters used in the model is defined in Table 2. Most values of the parameters are available from other studies. The new parameter introduced in the proposed system is not available. Therefore, these variables are obtained through curve fitting. There are two sources of data employed for the model fitting, which are the accumulated confirmed infected cases in the United States and the accumulated recovered cases in the United States. In this study, the number of detected infected people $\boldsymbol{D(t)}$ is accumulated and then fitted with the number of total confirmed cases. Since the recovered people $\boldsymbol{R}$ can't get infected again, the $\boldsymbol{R(t)}$ is accumulated with the time in the proposed model, which can be fitted with the number of total recovered cases. In order to calculate the system of the differential equation, the initial value of variables $\boldsymbol{S(t)}$, $\boldsymbol{E(t)}$, $\boldsymbol{I(t)}$, $\boldsymbol{A(t)}$, $\boldsymbol{D(t)}$, $\boldsymbol{D_H(t)}$, $\boldsymbol{D_D(t)}$, $\boldsymbol{R(t)}$ in the system should be defined. In the case study, $\boldsymbol{S(0)}$ is the total population of the United States which is 331.4 million (U.S. Census Bureau, 2020). $\boldsymbol{E(0)}$, $\boldsymbol{I(0)}$, and $\boldsymbol{A(0)}$ are unknown and need to be assumed. $\boldsymbol{D_H(0)}$, $\boldsymbol{D_D(0)}$, and $\boldsymbol{R(0)}$ are 0 at the beginning of the pandemic. $\boldsymbol{D(0)}$ is 1 at the start date of the simulation according to the data. Table 3 shows the fitted value of parameters with different assumed initial conditions. Figure 2 shows the curve-fitting results using two sets of initial values. With the higher initial value of exposed people $\boldsymbol{E(0)}$, the curve-fitting results are better.

Table 3: Fitted parameters based on the different initial values

	Trial 1	Trial 2
$S(0)$	331,400,000	331,400,000
$E(0)$	58,400	58,400
$A(0)$	300	200
$I(0)$	540	350
β_c	0.9865	0.9972
θ_I	0.6157	0.6829
θ_A	0.5076	0.244
γ_H	0.0046	0.0001759
γ_I	0.00385	0.00007195

Figure 2: Fitted results of two curves

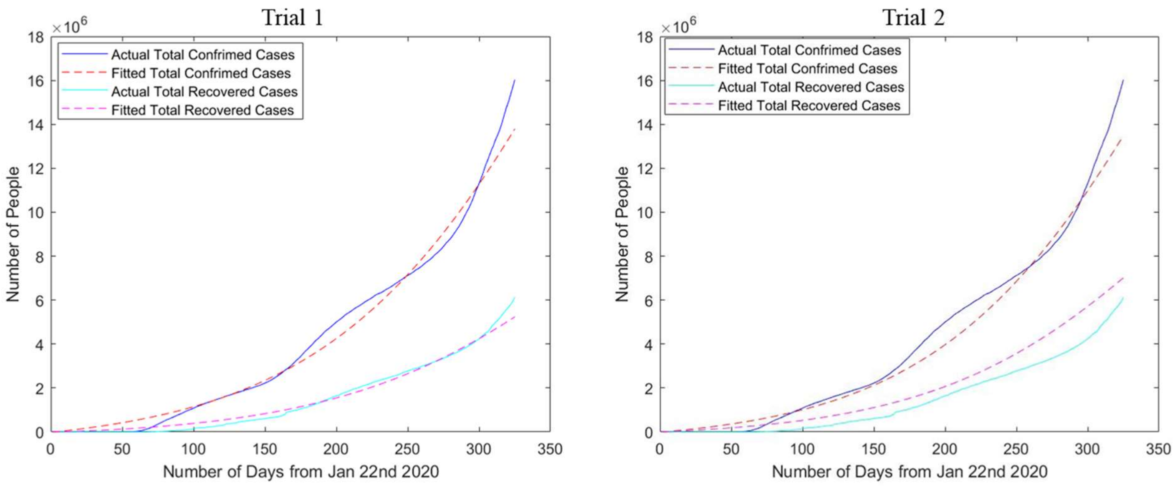


Figure 3: Contour plot of the basic reproduction number with $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$

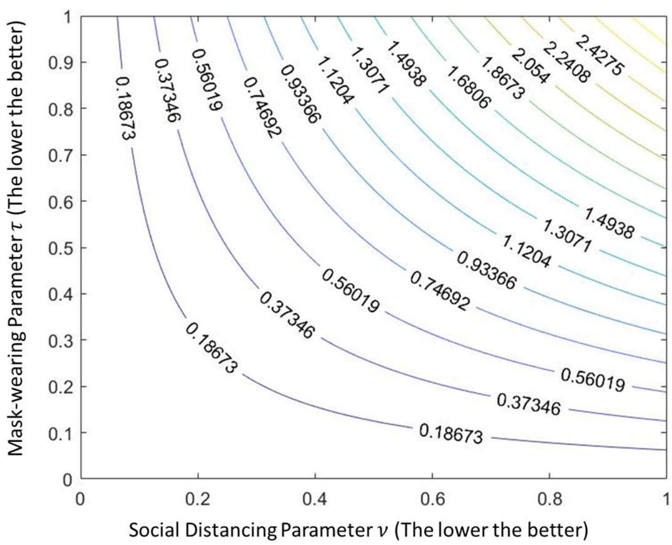
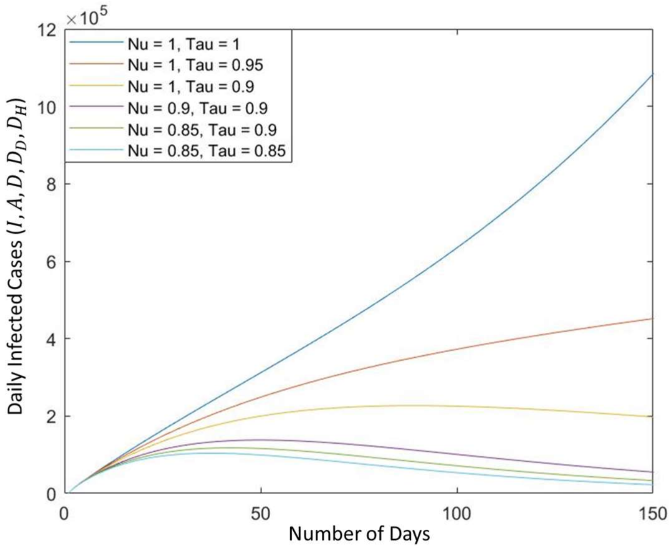


Figure 4: Number of infected cases with different $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$



The proposed model is able to simulate the COVID-19 pandemic for a certain time interval with limited known parameters. During the curve-fitting process, social distancing parameter $\boldsymbol{\nu}$ and mask-wearing level parameter $\boldsymbol{\tau}$ are equals to 1, which have no effects on the simulation. The change of $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$ can simulate the effect of different levels of response policies made by the government. It should be noted that, for both $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$, the bigger values mean the lower level of social distancing and mask-wearing. Figure 3 shows the contour plot of the basic reproduction number. The basic reproduction number becomes lower than 1 when $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$ are lower than 0.3. Figure 4 show the curve of all infected people $(\boldsymbol{I}(\boldsymbol{t}) + \boldsymbol{A}(\boldsymbol{t}) + \boldsymbol{D}(\boldsymbol{t}) + \boldsymbol{D_D}(\boldsymbol{t}) + \boldsymbol{D_H}(\boldsymbol{t}))$ with different levels of $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$. The predicted results show that the pandemic could be over within 150 days of 15% of people follows strict mask-wearing and social distancing policies at the early stage of the pandemic.

DISCUSSION AND CONCLUSIONS

The study case of the U.S. validates the feasibility of the proposed model. The model can predict and simulate the spread of the COVID-19 pandemic. Based on the case study, the COVID pandemic can be controlled and prevent by introducing the policy like social distancing and mask-wearing. The proposed model predicts the results that can reflect the effect of different levels of response policies. Based on the proposed model and other tools, it is possible to make policies that not only control the number of cases below a certain level but also can have limited effects on the economy. The future work of this study can focus on modifying the model for susceptible people that has multiple groups with different parameters. Other parameters in the proposed model can also be employed for determining the effects of other kinds of response policies in the early stage of pandemic.

REFERENCES

Bialek, S., Bowen, V., Chow, N., Curns, A., Gierke, R., Hall, A., Hughes, M., Pilishvili, T., Ritchey, M., Roguski, K., Silk, B., Skoff, T., Sundararaman, P., Ussery, E., Vasser, M., Whitham, H., & Wen, J. (2020). Geographic Differences in COVID-19 Cases, Deaths, and Incidence — United States, February 12th–April 7th, 2020. MMWR. Morbidity and Mortality Weekly Report, 69(15), 465–471. https://doi.org/10.15585/mmwr.mm6915e4

CDC. (2020, March 28th). COVID Data Tracker. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#trends_dailytrendscases

Centers for Disease Control and Prevention. (2021). *A Weekly Summary of U.S. COVID-19 Hospitalization Data.* U.S. Department of Health and Human Services. https://gis.cdc.gov/grasp/covidnet/COVID19_5.html

Centers for Disease Control and Prevention. (2021). Percentage of Confirmed COVID-19 encounters that died in the hospital from selected hospitals, by week. U.S. Department of Health and Human Services. https://www.cdc.gov/nchs/covid19/nhcs/hospital-mortality-by-week.htm

Diekmann, O., Heesterbeek, J., & Metz, J. (1990). On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology, 28(4). https://doi.org/10.1007/bf00178324

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases, 20(5), 533–534. https://doi.org/10.1016/s1473-3099(20)30120-1

Egger, M., Johnson, L., Althaus, C., Schöni, A., Salanti, G., Low, N., & Norris, S. L. (2017). Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. F1000Research, 6, 1584. https://doi.org/10.12688/f1000research.12367.1

KRR, G., KVR, M., SSP, P. R., & Casella, F. (2020). Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality. SSRN Electronic Journal. Published. https://doi.org/10.2139/ssrn.3560688

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., & Lessler, J. (2020). The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of internal medicine, 172(9), 577–582.

Li, M. Y., Graef, J. R., Wang, L., & Karsai, J. (1999). Global dynamics of a SEIR model with varying total population size. Mathematical Biosciences, 160(2), 191–213. https://doi.org/10.1016/s0025-5564(99)00030-9

Lotfi, M., Hamblin, M. R., & Rezaei, N. (2020). COVID-19: Transmission, prevention, and potential therapeutic opportunities. Clinica Chimica Acta, 508, 254–266. https://doi.org/10.1016/j.cca.2020.05.044

Mathieu, E., Ritchie, H., Ortiz-Ospina, E. et al. A global database of COVID-19 vaccinations. Nat Hum Behav (2021)

Okuonghae, D., & Omame, A. (2020). Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria. Chaos, Solitons & Fractals, 139, 110032. https://doi.org/10.1016/j.chaos.2020.110032

Oran, D. P., & Topol, E. J. (2020). Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. Annals of internal medicine, 173(5), 362–367.

Pellis, L., Ball, F., & Trapman, P. (2012). Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R0. Mathematical Biosciences, 235(1), 85–97. https://doi.org/10.1016/j.mbs.2011.10.009
Tang, B., Bragazzi, N. L., Li, Q., Tang, S., Xiao, Y., & Wu, J. (2020). An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infectious Disease Modelling, 5, 248–255. https://doi.org/10.1016/j.idm.2020.02.001

Tang, B., Wang, X., Li, Q., Bragazzi, N. L., Tang, S., Xiao, Y., & Wu, J. (2020). Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions. Journal of Clinical Medicine, 9(2), 462. https://doi.org/10.3390/jcm9020462

U.S. Census Bureau (2019). *Age and Sex Composition in the United States: 2019.* https://www.census.gov/data/tables/2019/demo/age-and-sex/2019-age-sex-composition.html

van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences, 180(1–2), 29–48. https://doi.org/10.1016/s0025-5564(02)00108-6
World Health Organization. (2020*). Listings of WHO's response to COVID-19.* https://www.who.int/news/item/29-06-2020-covidtimeline

Zaremba, A., Aharon, D. Y., Demir, E., Kizys, R., & Zawadka, D. (2021). COVID-19, government policy responses, and stock market liquidity around the world: A note. Research in International Business and Finance, 56, 101359. https://doi.org/10.1016/j.ribaf.2020.101359

DECISION SCIENCES INSTITUTE

Predicting IPO underperformance in Indian markets using Machine Learning techniques

Rachit Agrawal
SJMSOM, IIT Bombay
Email: rachit.agrawal@sjmsom.in

Usha Ananthakumar
SJMSOM, IIT Bombay
Email: usha@som.iitb.ac.in

ABSTRACT

Predicting the IPO short-term returns is a challenging task due to the involvement of many determinants. Empirical analysis and literature have shown the presence of IPO underpricing, and their research is dependent upon linear regression (GLM) models. In this study, we have utilized various machine learning techniques on classified data to predict IPO underpricing in the Indian markets. The results show that the gradient boosting model (GB) has performed significantly better than GLM models and other ensemble techniques. The variable importance measure indicates that profit margin, market sentiments, debt to assets ratio are the most important predictors of IPO short-term returns.

KEYWORDS: Initial Public Offerings, IPO underpricing, Ensemble techniques, Gradient boosting model, Prediction

INTRODUCTION

Financial markets are an integral part of any nation's economic growth engine and allow for the allocation and flow of funds from the low utility to high utility applications. The capital market helps to improve the allocation of funds from savers and investors with varying degree of needs and goals, distributed towards their most efficient uses to increase investment efficiency. This capital market's efficiency is a characteristic of the market in which it operates and relies on the flow of information for its speed, accuracy, and reliability to be useful for different stakeholders to form actionable insights and make informed decisions. The quality of these signals and their dissemination from the overall noise generated incessantly in the markets helps in price discovery mechanism and achieves this required efficiency in the financial markets to realize growth.

Initial Public Offering (IPO) is one of the processes, that contributes to the distribution of funds in capital markets. IPOs are part of a primary market where corporates raise the request for capital against equity shares to undertake new/expansion projects, from the investors willing to take the risk in the stated endeavor of the company with the prospect of earning higher returns. In this market, although, unlike the secondary market where price discovery happens by bringing together the buyers and sellers with supply-demand dynamics of the financial instruments, the bookmakers, i.e., investment banks and its consortiums, ascertain a price for these equity shares. Investment Banks based on their judgment of the company's shares demand at various price points about various intrinsic operational, financial, management-related factors deterministic to such process and capital needed by the company for the stated purpose. We primarily observe in emerging and developed economies that IPOs are mostly underpriced to attract high demand for such instruments. IPO's pricing differs significantly with varying degrees of underpricing and overpricing.

An IPO is said to be underpriced when the offer price of the issue is below the closing market price on the listing day. Whenever an issue is underpriced, it is gain for investors and loss of capital for issuing company and vice versa for overpricing of the issue. The understanding of determinants and driving factors for such underpriced and overpriced IPO is not properly understood yet. With this varying degree of pricing, investors can experience major losses in both the long term and the shorter term. In worst cases, total capital is lost; IPO may be delisted from the market.

The IPO market in India is on the boom, and a lot more companies are issuing equity shares in the capital market. Every year more than 30-35 companies on average raise capital from retail and institutional investors. This market has been improving over the past three years due to buoyant equity markets, increasing financial savings, and outperformance of recent IPO listings. However, not all the IPOs have generated positive returns for the investors. There have been instances of significant short-term losses and delisting in the IPO markets.

In this study, we have analyzed the underpricing patterns of historical IPOs to detect and predict the decisive variables in IPO listing day performance. The study could help retail investors identify important determinants for IPO underpricing in the market and protect investors from losing money in short-term investments. As the efficiency of a capital market is enhanced by the information and signal transmission to different stakeholders, such kind of exercise is with merit to attract investors to the market with a significant understanding of the market and gain confidence while investing in IPOs of Indian companies.

Machine learning under data science models can learn itself when a large enough data set is fed to the model to learn. This study utilizes classification techniques like logistic regression, random forest, bagging and gradient boosting, to predict the performance of IPOs.

LITERATURE REVIEW

(Stoll and Curley,1970; Ibboston, 1975) laid the foundation of IPO pricing and mispricing in accordance with short-term first-day returns. They observed that there was a significant price appreciation just after the equity offering issue is listed and initial gains in IPOs performance was exceptionally high.

Many theoretical studies have been carried out regarding the factors responsible for such a level of underpricing with IPOs. (Baron, 1982) attributed the underpricing of IPOs to two theories, first information asymmetry theory and second to the market uncertainty. According to the information asymmetry argument, the issuer has more idea about the operational and functional aspects of the business as they offer an incentive by underpricing that compensation. Second, due to market uncertainty, issuers rely on bankers' expertise on market conditions regarding issue's demand. (Ritter & Welch, 2002) reviewed the US market IPOs underperformance from 1980 to 2001, concluding that asymmetric information models are not the primary drivers of underperformance. They believed that agency conflicts and share allocation issues could explain the dramatic variations in underpricing.

Another theory by (Habib & Ljungqvist, 2001) tries to explain the voluntary underpricing of an issue by the issuer because they release a small portion of their holding to the public shareholders. (Loughran and Ritter, 2002) combine this with the fact that rather than maximizing their gain in the issue of fresh equity in isolation, they have their stake in maximizing their wealth. Hence, when an issuer releases a small portion of its shareholding for the public below fair market value, it acts as a strong incentive for investors and attracts a lot of demand from the public. By foregoing fractional gains on a small portion of holding, they attract higher valuation in the long term for their residual shareholdings.

Although most of the studies have been done for developed economies, (Loughran, Ritter, and Rydqvist,1994) finds underpricing in developing nations to be more than that observed in developed economies. (Gao, 2010; Tian, 2011) showed that asymmetric information theory does not work in a developing nation, like China. (Procianoy & Cigerza, 2007) used a multivariate linear regression model to explain the performance of IPO in emerging economies of Brazil, India, and China with determinants of performance as offer size, investment bank reputation, final offer price, market performance, goods produced with high tech content, GDP, FDI and interest rate and inflation of the country. Most of these theories have been subjected to rigorous empirical testing using firm-specific and market specific-factors. The empirical evidence presented in the literature is notably in favor of the asymmetric information theory.

Several researchers have analyzed IPO returns using a variety of computational intelligence techniques. (Luque, Quintana, and Isasi 2012) focus mainly on the offering characteristics to predict IPO returns using a genetic algorithm. (Robertson, Golden, Runger, and Wasil, 1998) construct an OLS regression and neural network models to predict the first day returns of IPOs, the empirical findings of their study show that the predictions produced by neural network models were better than predictions produced by OLS regression. (Quintana, Saez, & Isasi, 2017) have tried to explain the IPO underperformance in US markets using machine learning models and have concluded that random forests outperform the eight popular machine learning algorithms: instance-based learning algorithms; least median of squares regression; locally-weighted learning; M5 model trees; M5 model rules; multilayer perceptron; radial basis function networks; and support vector machines trained with sequential minimal optimization in terms of mean and median predictive accuracy. (Baba & Sevil, 2020) studied IPO returns in the Istanbul equity market and have got similar conclusions regarding the random forest algorithm.

Considering Indian markets, (Krishnamurti & Kumar, 2002) tested the underpricing for the time lag between final allotment and listing and found that it is an important factor in underpricing. The reason stated for the relation between the two was that the investors perceived time lag as a risk and required additional compensation for the risks involved. (Ghosh, 2002) studied the relationship between uncertainty and age of the firm as determining factors of underpricing and found a positive relationship between uncertainty risk and underpricing but the firm’s age could not explain the initial returns. (Bansal & Khost, Anna, 2012) analyzed Indian IPOs after the global financial crisis and found that firm's age, IPO years, book building pricing mechanism, ownership structure, issue size, and market capitalization explained 44% of the variation in issuer underpricing. (Chhabra, Kiran and Sah, 2017a) find the variables that signal information are highly significant and companies with high information disclosure experience less underpricing. On the other hand, (Chhabra, Kiran, Sah and Sharma, 2017b) find the informational variables less effective in explaining the IPO underpricing.

The present literature on Indian capital markets is constricted to regression models with smaller data sets, limited variables, and more focused on long-term capital gains. We plan to analyze the Indian equity markets for short-term IPO returns using the machine learning techniques from 2007-2020 (pre-COVID) which includes listings from a period including the pre-global financial crisis to pre-COVID 19.

SAMPLE DATA

Data

The data set used in this study consists of first-day trading returns of 262 public offerings listed on NSE during the period 2007 to 2020 (pre-COVID) with a minimum post-issue paid-up capital of Rs 10 crore. Table I shows the IPO proceeds during this period. In this period, a total of 429 major firms went public, raising around $42 billion of capital. The volume of IPO's peaked in 2007 with 108 deals. However, we see a significant dip in the interest post-2007. This reduction in IPO activity was largely driven by the rapid decline in Indian stock exchanges following the breakout of the global financial crisis in the third quarter of 2008. The IPO activities bounced back in 2009-10. However, IPO activities took a sudden dip in the years 2013 and 2014 due to poor sentiments, volatile secondary markets, and promoters not getting the right valuations. In the following years, especially since 2015, the global favorable monetary conditions spurred economic growth and sectoral developments in many emerging markets. The year 2017 remains the most productive year for capital raising, with the public injecting ~$11 billion into equities. Due to data inconsistencies, we removed 167 firms from our dataset. In our sample of 262 IPOs from 2007 to 2020, the average underpricing is 15.7 percent. Approximately 63 percent of the IPOs end the first day of trading at a closing price greater than the offer price and about 37 percent have a first-day zero or negative returns. The IPO data used in the empirical analysis is obtained from the NSE website and CMIE prime database.

Table I. IPO Proceeds

Year	*Number of IPO*	*IPO Proceeds (in US$bn)*	*% Percentage*
2007	108	4.85	12%
2008	39	2.62	6%
2009	22	2.76	7%
2010	66	5.19	12%
2011	40	0.85	2%
2012	13	0.98	2%
2013	5	0.18	0%
2014	7	0.17	0%
2015	21	1.93	5%
2016	27	3.79	9%
2017	38	10.75	26%
2018	25	4.53	11%
2019	16	1.81	4%
2020 (pre-covid)	2	1.48	4%
Total	**429**	**41.9**	**100%**

Source: NSE India

Variables

The IPO process is closely related to the firm qualities, market timing, agency issues, investors' interest, asymmetric information, and so on. For this study, we have selected variables based on previous IPO underpricing literature and some new firm characteristics variables. The description of the variables selected and descriptive statistics of the variables are provided in Table II. For market characteristics, we will be using broader market returns ("BSE") as a proxy for market sentiments. For firm characterstics, we have used Pre-IPO year data as the basis for calculation of ratios and margins.

To predict the IPO underpricing, we first select the commonly used measure of underpricing, which is expressed as

$$U_i = \frac{P_{oi} - P_{ci}}{P_{oi}} \quad (1)$$

where,

U_i is the degree of underpricing

P_{oi} represents the offered price for stock i

P_{ci} is the closing price for stock i on the listing day

The negative values of U_i represent underpriced IPO while the positive values suggest an overpriced IPO. U_i is the dependent variable for our study. We have classified the dependent variable as “underpriced “ and “overpriced” for further analysis.

Table II. Variables & descriptive statistics

Variables	*Explanation*	*Median*	*Mean*	*Std. Dev*
Firm Characteristics (Calculated based on Pre-IPO year ending financials)				
Profit Margin (“PM”)	Firm net profit divided by the revenues of the business	8.4%	8.3%	0.19
Asset Turnover Ratio	Asset turnover ratio measures the value of a company's sales or revenues relative to the value of its assets.	0.77	0.93	0.75
Return on assets (“ROA”)	Return to assets ratio measures the profit of the company relative to its assets	0.07	0.07	0.06
Debt to Asset ratio (“D/A ratio”)	The ratio of company's debt to its assets	0.22	0.26	0.19
Firm Age	The age of the company since the date of incorporation till IPO date	15.0	17.9	14.9
Total Assets (Rs crores)	Total assets of the company, an indicator of company size	5,054	33,023	-
Offering Characteristics				
Company sold (%)	IPO size as the ratio of the company's total market cap	25.5%	26.7%	0.11
Width of the price range (RANGE)	The difference between the offering price ranges	10.0	15.6	16.40
Equity Offer Size (Rs mm)	The amount of capital raised by the company (Issue size), indicative of the size of the IPO	19,123	61,341	-
Issue expenses (as % of IPO size)	The IPO expenses in million rupees, as outlined in the final prospectus	6.7%	6.8%	0.02
Market Characteristics				
BSE 500 3M return ("market sentiment”)	3-months BSE 500 return before the listing date of IPO as a proxy for market sentiments	2.4%	2.9%	0.11

RESEARCH METHODOLOGY

We intend to utilize classification algorithms to predict IPO underpricing and compare the techniques based on their predictive ability. We also aim to identify variables that are significantly better predictors for IPO underpricing. Accordingly, we selected the following set of methods to analyze our data - logistic regression, decision trees and ensemble techniques like bagging, random forest, gradient boosting.

Predictive Models

Logistic Regression

Logistic regression (LR) is a statistical method that predicts the probability of a categorical response for a given set of independent variables. (Hosmer & Lemeshow, 1989) provide a comprehensive introduction to logistic regression analysis.

The logistic regression model uses the odds ratio, which represents the probability of an event of interest compared with the probability of not having an event of interest. The model is based on the natural logarithm of the odds ratio given by

$$\ln(p/1\text{-}p) = \beta_0 + \beta_1.x_1 + \beta_2.x_2 + \cdots + \beta_{k.}\, x_k \quad (2)$$

where p is the probability of success, p/1-p is the odds ratio, and β_0, β_1, β_2, … β_k are the parameters of the model. LR uses maximum likelihood method to estimate the model parameters.

Decision Trees

Decision trees are one of the most used predictive modeling algorithms in practice. Decision trees were first applied to language modeling by (Bahl et al.,1987) to estimate the probability of spoken words. A decision tree is a predictive model, which is a mapping from observations about an item to conclusions about its target value. Decision trees work by doing successive binary splits. The first split will yield the biggest separation or distinction in two groups of data. Each subgroup is then split until some stopping criteria is reached. In the tree structures, leaves represent classifications (also referred to as labels), non-leaf nodes are features, and branches represent features that lead to the classifications.

Bagging or Bootstrap Aggregating

Bagging (Breiman, 1996) is a method for fitting multiple versions of a prediction model and then combining them into an aggregated prediction. In the bagging algorithm, bootstrap copies of the original training data are created, the regression or classification algorithm (commonly referred to as the base learner) is applied to each bootstrap sample and in the classification context, new predictions are made by taking the majority vote prediction for the classes from across the predictions made by the decision trees. Bagging effectively reduces the variance of an individual base learner because of the aggregation process.

Random Forest

Random Forest (Freund and Schapire, 1996) combines the two concepts of bagging and random selection of features by generating a set of classification trees where the training set for each tree is selected using bootstrapping from the original sample set and the features considered for partitioning at each node is a random subset of the original set of features. Random Forest has become a commonly used tool due to its ability to handle large features with small samples and improved accuracy. The random sampling and bootstrapping together reduces the correlation between the generated trees and majority voting of the class responses reduce the variance of the error, providing an improvement over bagging algorithm.

When building decision trees, each time a split in the tree is considered, a random selection of m predictors is chosen as a subset of split candidates from the full set of predictors. The number of predictors considered at each split (m) is approximately equal to the square root of the total number of predictors, p.

The random forest trees are insensitive to skewed distributions, outliers, and missing values; they are considered as one of the most efficient predictive ML techniques.

Gradient Boosting Model

Gradient Boosting (GB) (Freund & Schapire, 1997; Friedman et al., 2000) is a machine-learning algorithm for regression and classification problems that combines the output of many weak predictive models to produce a final robust predictive model.

GB procedures are invariant under all monotonic transformations of a single input variable (e.g., logarithm transform) and are not sensitive to outliers (this algorithm isolates outliers only in separate nodes without affecting the performance of the final model). GB is insensitive to multicollinearity, and it is more robust due to better handling of uncorrelated inputs (Friedman, 2001). GB can select and rank the variables, which provides a feasible way to compare IPO underpricing predictors.

The statistical framework describes boosting as a numerical optimization problem whose goal is to minimize the loss of the model by adding weak learners using gradient descent-like process. The weights of the selection of the training data set for the latter learners are not equal. Samples with larger errors have higher weights. The latter learners in GB are adjusted based on the errors made by the previous learners. GB has 3 major parts -:

Loss Function - The role of the loss function is to estimate how good the model is at making predictions with the given data.

Weak Learner - A weak learner is one that classifies our data but does poorly. It has a high error rate. Decision tree models are often selected as weak learners.

Additive Model - This is the iterative and sequential approach of adding the trees (weak learners) one step at a time. Each iteration should reduce the value of our loss function.

Boosted trees are grown sequentially; each tree is grown using information from previously grown trees to improve performance. By fitting each tree in the sequence to the previous tree's residuals, we are allowing each new tree in the sequence to focus on the previous tree's mistakes and thus do better.

Evaluation Metrics

We have analyzed the above methods based on their predictive accuracies, sensitivity, AUC, and RIS. Sensitivity plays an important role in the study as the Type 1 error is a more important component for results than the Type 2 error.

- Type I error (Positive class – “Overpriced”): If the model predicts an overpriced IPO as underpriced, it could incur a significant loss to an investor. Higher the sensitivity lower would be the Type I error
- Type II error: If the model predicts an underpriced IPO as overpriced, the investor would not be incurring losses as the investor would not invest money in the IPO based on the model’s predictions.

AUC stands for Area under the ROC Curve. ROC curve or Receiver Operating Characteristic curve is a graph that shows the performance of a classification model at all classification thresholds. AUC provides an aggregate measure of performance across all possible classification thresholds.

Relative influence score (RIS) is a measure of how useful a particular variable is to a model by quantifying the importance of the variable marginalizing on other variables. (Friedman & Meulman, 2003), showed that measures are based on the number of times a variable is selected for improving the model. RIS score is used to study variable importance and its effect on the prediction model.

Table III. Model Comparison

Method	***Comparative results and hyperparameters***
Decision Trees	The test data accuracy for the pruned tree is 0.58. The tree predicted BSE 3M return, profit margin and asset turnover ratio as the nodes.
Logistic Regression	The regression model has 0.59 accuracy. The model predicted BSE 3M return as the significant variable
Bagging	• Full tree model with 1500 individual base learners • The bagged tree has an accuracy of 0.62 • The model predicted BSE 3M return and profit margin as the most important variables
Random Forest	• Full tree model with 1000 individual base learners and 4 variables randomly sampled for each split • The random forest model has an accuracy of 0.64 • The model predicted profit margin and BSE 3M return as the most important variables
Gradient Boosting Model	• GB Model with the optimized hyperparameters o Number of Trees: 3000 o Learning Rate: 0.1 o Cross validation folds: 6 o Tree Depth: 2 • The model has an accuracy of 0.68 higher than the bagged and random forest model • The model predicted include BSE 3M return, D/A ratio, profit margin, asset turnover ratio and % of company sold as the most important variables

EMPIRICAL RESULTS

We applied the various classification methods on a training dataset which was assigned 70% of the observations randomly using stratified sampling. On the remaining 30% sample, we tested our models. Table III summarizes the outcomes obtained on the application of models to the testing dataset.

Table IV summarizes the predictive performances of the gradient boosting (GB) model, logistic regression (GLM), pruned tree, random forest, and bagging. The test sample has a significant higher AUC (also shown in Figure I), higher accuracy, and higher sensitivity on the GB model, which indicates that the GB model has overall better prediction performance compared to other models.

Figure I. AUC Comparison

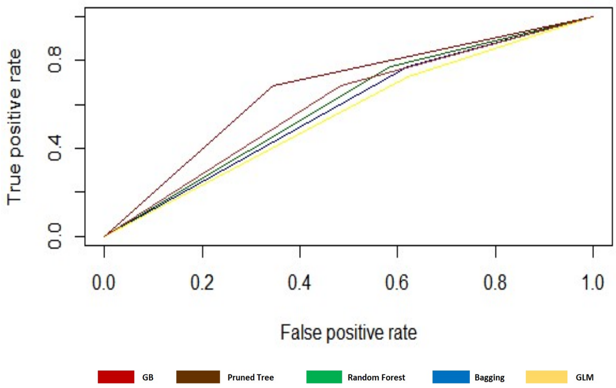
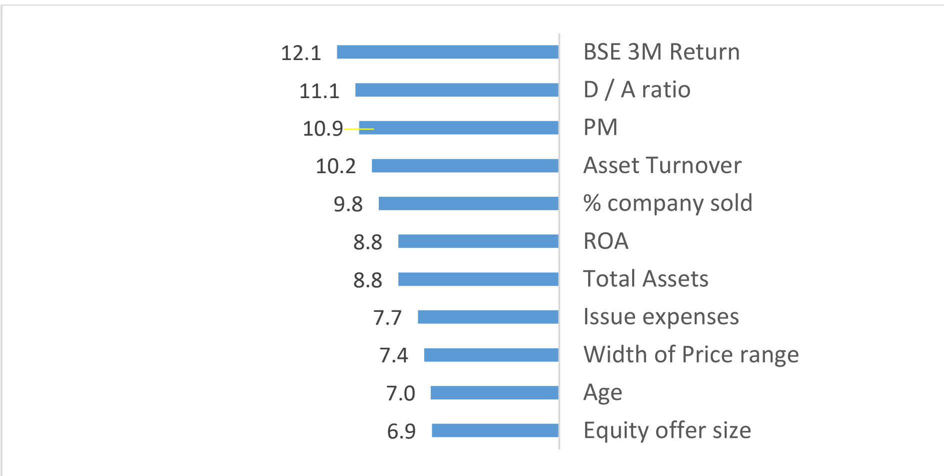


Table IV. Model Evaluations & Results

Models	***AUC***	***Accuracy***	***Sensitivity***
Pruned Tree	0.60	0.58	0.51
GLM	0.55	0.60	0.38
Bagging	0.58	0.62	0.38
Random Forest	0.59	0.64	0.41
GB	0.67	0.68	0.65

Figure II. Variable Importance (GB)



Based on the variable importance graph of the GB model (as shown in Figure II), we find that BSE 3M return, D/A ratio, profit margin, asset turnover ratios, are the most important indicators of an IPO's underpricing. The study also highlights those firm characteristics have higher variable importance than other listing characteristics.

CONCLUSION

Due to the involvement of many determinants with very different explanatory power, the presence of outliers, and data inconsistencies, predicting IPO underpricing has been a challenging task. We selected 11 determinants based on numerous empirical findings of previous studies. The outcomes of this study show that profit margin, market short -term sentiments and leverage ratio i.e. D/A ratio are the most important predictors for IPO underpricing in the Indian markets. Based on the results, it can be said that firm characteristics like margin measures, leverage ratios and asset quality, market characteristics like short-term market sentiment and listing characteristics like issue expenses should be considered for investing in an IPO for a short-term capital gain perspective.

The gradient boosting model in predicting short term gains maybe of relevance as both IPO issuers and investors are highly concerned with the uncertainty and market response regarding IPO price and performance. This study shows that the gradient boosting model has advantages in predicting IPO underpricing over GLM models in three aspects: adaptability to large numbers of input variables, it provides a rank (RIS) of these inputs based on their contribution to the prediction, and it can properly handle predictors that show large skewness, kurtosis, and violate GLM assumptions.

The empirical findings of this study add to the existing literature by emphasizing the accuracy and methodological advantages of the ensemble techniques in predicting the IPO underpricing.

REFERENCES

[1] Baba, B., & Sevil, G. (2020). Predicting IPO initial returns using random forest. *Borsa Istanbul Review*, 20(1), 13-23.

[2] Bansal R., & Khanna, R. (2012). IPOs underpricing and money “left on the table” in Indian market. *International Journal of Research in Management, Economics and Commerce*, 2, 106-120.

[3] Baron, D. (1982) A Model of the Demand for Investment Banking Advice and Distribution Services for New Issues. *Journal of Finance*, 37, 955-976.

[4] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

[5] Chhabra, S., Kiran, R., & Sah, A. N. (2017a). Information asymmetry leads to underpricing: Validation through SEM for Indian IPOs. *Program*, 51(2),116-131.

[6] Chhabra, S., Kiran, R., Sah, A. N., & Sharma, V. (2017b). Information and performance optimization: A study of Indian IPOs during 2005-2012. *Program*, 51(4), 458-471.

[7] Desai, V. S., & Bharati, R. (1998). A comparison of linear regression and neural network methods for predicting excess returns on large stocks. *Annals of Operations Research*, 78(0), 127-163.

[8] Freund, Y., & Schapire, R.E. (1996). Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference*, 13,148-156.

[9] Freund, Y., & Schapire, R.E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.

[10] Friedman, Jerome. (2000). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 1189-1232.

[11] Friedman, Jerome & Meulman, Jacqueline. (2003). Multiple additive regression trees with application in Epidemiology. *Statistics in medicine*, 22, 1365-81

[12] Gao, Y. (2010). What comprises IPO initial returns: Evidence from the Chinese market. *Pacific-Basin Finance Journal*, 18(1), 77-89.

[13] Ghosh, Saurabh (2002). Underpricing of Ipos: The Indian Experience Over the Last Decade.*SSRN Electric Journal.*

[14] Habib, M.A. and A.P. Ljungqvist (2001). Underpricing and Entrepreneurial Wealth Losses in IPOs: Theory and Evidence. *The Review of Financial Studies*, 14(2), 433-458.

[15] Hosmer, D., & Lemeshow, S. (1989). Applied logistic regression. New York: Wiley.

[16] Krishnamurti, C. and Kumar, P. (2002). The initial listing performance of Indian IPOs, *Managerial Finance*, 28 (2), 39-51.

[17] Lin, C., & Hsu, S. (2008). Determinants of the initial IPO performance: Evidence from Hong Kong and Taiwan. *Applied Financial Economics*, 18(12), 955-963.

[18] Ljungqvist, A., & Wilhelm, W. J. (2003). IPO Pricing in the Dot-Com Bubble. *The Journal of Finance*, 58(2), 723–752.

[19] Loughran, T., Ritter, J. R., & Rydqvist, K. (1994). Initial public offerings: International insights.*Pacific-Basin Finance Journal*, 2(2-3), 165-199.

[20] Loughran, T. and J.R. Ritter, 2002, Why Don't Issuers Get Upset About Leaving Money on the Table in IPOs?. *Review of Financial Studies,* 15, 413-443.

[21] Luque, C., Quintana, D., & Isasi, P. (2012). Predicting IPO underpricing with genetic algorithms. *International Journal of Artificial Intelligence*, 8(S12), 133-146.

[22] Marshal, B. B. (2004). The effect of firm financial characteristics and the availability of alternative finance on IPO underpricing. *Journal of Eco-nomics and Finance*, 28(1), 88-103.

[23] Pande, A., & Vaidyanathan, R. (2009). Determinants of IPO underpricing in the National Stock Exchange of India. *ICFAI Journal of Applied Finance*, 25(1), 84-107.

[24] Procianoy, Jairo Laser & Cigerza, Gilles. (2007). IPOs in Emerging Markets: A Comparison of Brazil, India and China. *SSRN Electronic Journal*.

[25] Pande, Alok & Vaidyanathan, R. (2007). Determinants of IPO Underpricingin the National Stock Exchange of India. *IUP Journal of Applied Financ*e, 15(1), 17-18.

[26] Quintana, D., Saez, Y., & Isasi, P. (2017). Random forest prediction of IPO underpricing. *Applied Sciences*, 6(7), 63.

[27] Ritter, J. R. (1984). The "hot issue" market of 1980. *Journal of Business*, 57(2), 215-240.

[28] Ritter, J. R., & Welch, I. (2002). A review of IPO activity, pricing, and allocations. *The Journal of Finance*, 57(4), 1795-1828.

[29] Robertson, S. J., Golden, B. L., Runger, G. C., & Wasil, E. E. (1998). Neural network models for initial public offerings. *Neurocomputing*, 18(3), 165-182.

[30] Stoll, H.R. and A.J. Curley. 1970. Small Business and the New Issues Market for Equities. *Journal of Financial and Quantitative Analysis*, 5(3), 309-322.

[31] Tian, L. (2011). Regulatory underpricing: Determinants of Chinese extreme IPO returns. *Journal of Empirical Finance*, 18(1), 78-90.

DECISION SCIENCES INSTITUTE
Herding Behavior and Its' Effect on Portfolio Returns

Joshua Demik, Roosevelt University, jdemik@mail.roosevelt.edu

Asghar Sabbaghi, Ph.D., Roosevelt University, asabbaghi@roosevelt.edu

ABSTRACT

This study will examine the phenomenon of herding and its impact on the security pricing and overall market return as well as individual returns. Herding is known as an excessive irrational tendency of investors ignoring the fundamental information presented to them that act together in the markets (Gębka & Wohar 2013). Herding can directly lead to the security market distress and to the increase in trading volatility. In this study, we will examine how the security market under this irrational tendency would directly affect the prices of securities, and returns on individual's portfolios.

Keywords: Behavioral Finance, Herding behavior, Cross-Sectional Standard Deviation (CSSD), Cross-sectional absolute deviation (CSAD).

INTRODUCTION

Herding behavior has received much attention in finance and security market, where a group of investors tends to trade in the same direction over a period, leading to observed behavior patterns that are correlated across individuals (Bikhchandani et al., 1992). In this study, we will examine the financial phenomenon of herding and its impact on the security pricing and overall market return as well as individual returns. Herding is generally known as an excessive irrational tendency of investors ignoring the fundamental information presented to them that would act together in the markets (Gębka & Wohar 2013). Herding can directly lead to market distress and led to the increase in trading volatility.

There is a distinction between two different types of herding: (1) spurious, or unintentional, herding which is commonly a result of publicly known news being interpreted similarly and causing efficient price corrections; but this is not usually a cause of excess volatility in the markets. (2) Intentional herding, when investors possess knowledge of other investors' decisions and knowledge and this in turn affects their investment decisions. The securities market is a very large and complex market that is traded nearly every day on a domestic and international level. There is a constant flow of information to each and every investor that comes from different sources, ideologies, and credibility. It is human nature to take what we are given and interpret it in our own light. This very nature is what led to Behavioral Finance and the study of its effects on the market.

Christie and Roger D. Huang (1995), argue that in a market setting, herds are characterized by individuals who suppress their own beliefs and base their investment decisions solely on the collective actions of the market, even when they disagree with its predictions. Accordingly, they believe that traditional herd exist when investors are drawn to the consensus of the market, implying that individual returns would move in harmony with the market return. According to Igual and Santamaria (2017) one of the main theories that dominate the Behavioral Finance field is the irrational phenomena of herding. It is the intention of this study to examine how the market under this irrational tendency would directly affect the prices of securities, and returns on individual's portfolios. More specifically, in this study, we plan to analyze and answer the following questions: (1) Is the return dispersion of individual assets from the market return a significant indicator of herding in financial markets? And (2) What affects does herding have on security pricing and overall market return as well as individual returns?

We believe that the market is primarily driven by emotion, at least on a retail level. According to the literature, the average investors have difficult time removing their emotions from their investing activity (Qawi 2010). This issue has been studied extensively in the literature and received much attention. In this study, we plan to analyze a number of deeper implications that would bring light to the direct effects of the individual, as well as the broader market performance. More specifically, we examine the relationship between the market return and the individual stock returns in the context of herding, and highlight the degree, if any, of herding that has taken place over the past twenty-year period of stock returns.

In this study, we will follow the model from Dang and Lin (2016) which had been recreated from an empirical study of herding from Christie and Huang (1995) as well as Cheng et al. (2000). However, in this study, we will focus on the Dow Jones Industrial Average over a twenty-year period from 2000 until 2020. First, we will review the theoretical approaches of herding in the literature, as well as, the empirical studies involving the cross-sectional return dispersion of assets. Next, we will define our analytical framework and the data to develop our model. Finally, we will summarize the empirical findings, and conclusions about herding in the market.

LITERATURE REVIEW

The phenomena of social herding in the markets has been extensively studied in the literature from different perspectives. In Finance, particularly, herding has been viewed as a concern to market participants, as it contributes to the market inefficiency (Fama, 1970; Christie & Huang, 1995), and to the financial market volatility and instability (Bikhchandani & Sharma, 2001). Qawi (2010) has studied herding in finance and argues that everybody comes with this sort of 'bagage'. That implies that we all have our own inherent biases and heuristics.

There are two main avenues of thought in finance: (1) The Neo-classical theory of the rational investor who only makes decisions based on available data and proven mathematical theories. However, Qawi (2010) states that this method consists of incomplete information and abbreviated assumptions of reality that do not take into account any bit of human behavior element. This leads to the newer theory of psychological finance. It involves understanding market anomalies and accounting for these inherent biases. Qawi (2010) refers to the study by Pretcher (2001) on human behavior and its' reaction to market performance. Herding, the basic human function, is a response to the actions of others and stems from impulsive mental activity.

Pretcher (2001) concludes that herding is rooted within the limbic system of the human brain and it is immutable and impulsive. Qawi (2010) highlights another theoretical framework in regards to herding, called the 'Affect Pricing Model'. It works with the concept that individuals have perceptions of objects being valued as 'good' or 'bad', and that this mindset can be applied to individual stocks. He refers to a study by Kahneman (2002) that highlights another heuristic called the 'affect heuristic'. Gebka and Wohar (2013) argue that on an international level, there is not much herding to be found in the broader sense of the market. However, there are some herding tendencies found when you dig deeper into the returns of individual sectors and get rid of the broad index returns. They conclude that Fama's Efficient Market Hypothesis (EMH) holds true and is one main reason behind social herding tendencies within the market. The reason, they argue, is that the EMH states that the market is immediately receptive to the development of new information. So, the new information of the market is being accounted for immediately at the same time. As humans, it is in our nature to have the tendency to follow the majority, and that is what is occurring through the EMH. Chang et al. (2000) have studied herding tendencies across multiple international markets. They conducted the study in the U.S., Japanese, Hong

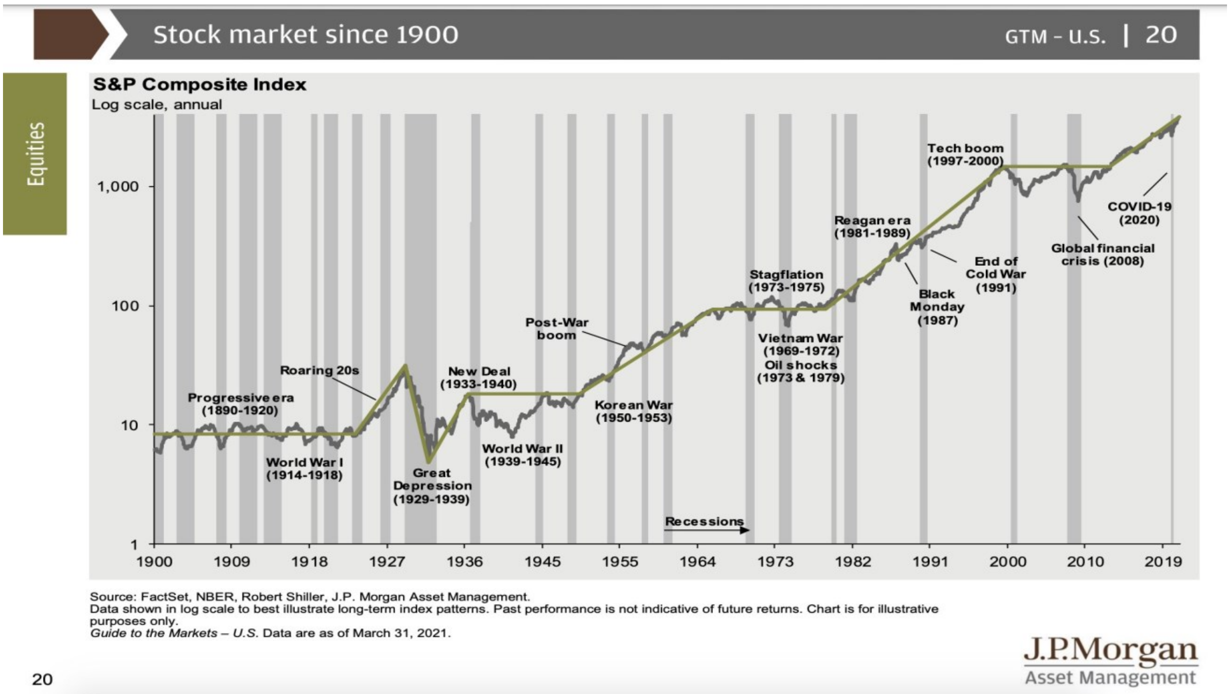
Kong, South Korean, and Taiwanese markets. They have used a variant of the empirical model from Christie and Huang (1995) and examined the investment behavior of market participants within different international markets (i.e., US, Hong Kong, Japan, South Korea, and Taiwan), with regard to their tendency to exhibit herd behavior. Their results for the US were consistent with those results reported by Christie and Huang (1995). They found no evidence of herding on the part of market participants in the US and Hong Kong and partial evidence of herding in Japan. However, they found significant evidence of herding for South Korea and Taiwan, the two emerging markets.

In their empirical analysis, Christie and Huang (1995) have focused on the price implications of herding by using cross-sectional standard deviation of returns. Therefore, in their study, herding behavior will be measured by decreasing dispersion of returns when individual returns changes are in harmony with market return. Because, there is herd behavior when individual returns follow the lead of the portfolio returns. They have also examined the herding behavior during the periods of abnormally large average price movements, or so-called market stress and conclude that, in contrast, the herding of individual returns around the market translates into a reduced level of dispersion. Therefore, during the market stress their conclusion supports the predictions of rational asset pricing models and suggests that herding is not an important factor in determining equity returns during periods of market stress. Thus, they had found that it was more common for prices to deviate from the market price in these times of crisis rather than follow it.

MODELING FRAMEWORK

In this section, we formulate herding behavior to capture the magnitude of the dispersion and apply it to long-term portfolio investing, and answer the question of whether or not herding will affect the long-term investors investment decisions. We use the model to describe the market volatility for the long-term investor. We use data from early 2000 when the financial tech bubble collapsed, and this gave the market a fresh valuation of asset pricing. The new age of technology has also gave investors more access to information and more insight to investment decisions made by the institutional investors. The data set would include the data from the Global Financial Crisis struck and scorned investors' returns and their market behavior.

We have used the daily price activity of the securities comprising the Dow Jones Industrial average from Yahoo!Finance, over a twenty year period spanning from 2000 until 2020. The daily price returns have been calculated by dividing the current stock price of the day by the lagged-one day price. The DJIA also captures the 30 largest stocks on the New York Stock Exchange (NYSE). This study considers price as being the determinant of return dispersion, and the DJIA's criteria for inclusion in the index is the 30 largest stocks, measured by price. There are skeptics on the rationale behind this, for the S&P 500 is comprised of the 500 largest stocks based on market capitalization and weighted on this criterion. However, the measurement of price will be sufficient for this study due to the main focus of this study revolving around price itself. There is one caveat in this data set, and that is that two companies from the DJIA had to be dropped from the individual securities list in order to maintain the consistency and continuity within the data set. The securities that have been removed from the data set are Visa (V) and Salesforce (CRM).



Our goal in this study is to examine the magnitude of the effect between the daily returns of the overall market index and the dispersion of returns the individual securities that comprise the index in comparison to the markets' daily returns. Furthermore, we aim to look into the effect herding will have on portfolio returns based off the calculation of securities dispersions on a daily level. The question to be answered by this is whether or not an individual investor with a buy and hold, long-term investment strategy will need to calculate the risk of current market volatility inherited from herding to better allocate their portfolio.

We assume that herding as a behavioral aspect of finance, will occurs when security return data mirrors the overall market or index returns. Investors will begin to ignore the information they have and stay away from their beliefs in order to make the investment decisions they think are best for them, by mimicking the investment decisions of other institutional investors. In this study, we follow Christie & Huang (1995), Cheng et al. (2000), and Dang & Lin (2016) to measure the herding index of the DJIA. More specifically, we use the Cross-Sectional Standard Deviation (CSSD) and the Cross-sectional absolute deviation (CSAD) to measure the absolute return dispersion of each individual market security that makes up the DJIA. The formulas are as follows:

$$CSAD = 1/N\sum |R_{i,t} - R_{m,t}| \quad (1)$$

$$CSSD = \sqrt{\frac{1}{N-1}\sum(R_{it} - R_{mt})^2} \quad (2)$$

Where $R_{i,t}$ is the stock return of a specific firm at time t, and $R_{m,t}$ is the cross-sectional average of all N stock returns that comprise the market portfolio at time t. Each of these calculations measure the average deviation of the individual stock returns from the market return ($R_{i,t}$ - $R_{m,t}$). In order to prevent any positive and negative values from canceling each other out, CSAD in (1) takes the absolute value of the deviation values while in equation (2), the CSSD squares the deviation values (Dang & Lin 2016). When looking at the values of deviation, any increasing values of CSAD or CSSD would imply that individual stock returns deviate from the market return, and this would imply that there is no evidence of herding between the individual returns and the market return. On the other hand, smaller values of deviation would imply the potential for herding, however, that may be spurious or intentional, yet to be known.

Furthermore, as Dang and Lin (2016) stated, different stocks may react differently to the market changes with different degrees of sensitivity, and would cause the dispersion to increase as the market return increases in absolute term. However, the presence of herd mentality during the extreme market movements will draw individual stock returns closer to the market return, and reducing dispersion.

In order to capture the herd mentality during extreme market movements, following Christie and Huang (1995), we have used dummy variable regression to test for herd behavior:

$$CSSD = \alpha + \beta_L D^L_t + \beta_U D^U_t + \varepsilon_t \qquad (3)$$

Where the two dummy variables D^L_t and D^U_t, assume a binary values: 0 when the market return on day t is outside of the upper or lower tails of the distribution, and a value of 1 if it does fall within the extreme upper and lower tails of the distribution. The criteria for the two dummy variables are the lower 5% , measured as -1.83% and lower of the distribution, and the upper 95%, measured as 1.7% and higher of the distribution. Thus, the lower tail cutoff has a value of -1.83% and the upper tail cutoff has a value of 1.7%. Any negative values of β_L or β_U that are statistically significant will indicate that there is a potential for herd behavior in the market (Dang & Lin 2016). We can recalculate this equation similarly for CSAD, by replacing CSSD by CSAD replaces CSSD in the equation.

Furthermore, according to Cheng et. Al (2000), in order for herding to exist, during larger market movements, there would need to be a non-linear relationship between $CSAD_t$ and $R_{m,t}$. They propose the following quadratic regression:

$$CSAD_t = y_0 + y_1|R_{m,t}| + y_2 R^2_{m,t} + \varepsilon_t \qquad (4)$$

Where $|R_{m,t}|$ is the absolute value of the market return at time t, and $R^2_{m,t}$ is the quadratic function of the market return at time t. The market would be exhibiting herd behavior if and when the dispersions increases at a less-than proportional rate or, decreases with the market return. This would result in a negative value and statistically significant value for the quadratic term (y_2) (Dang & Lin 2016).

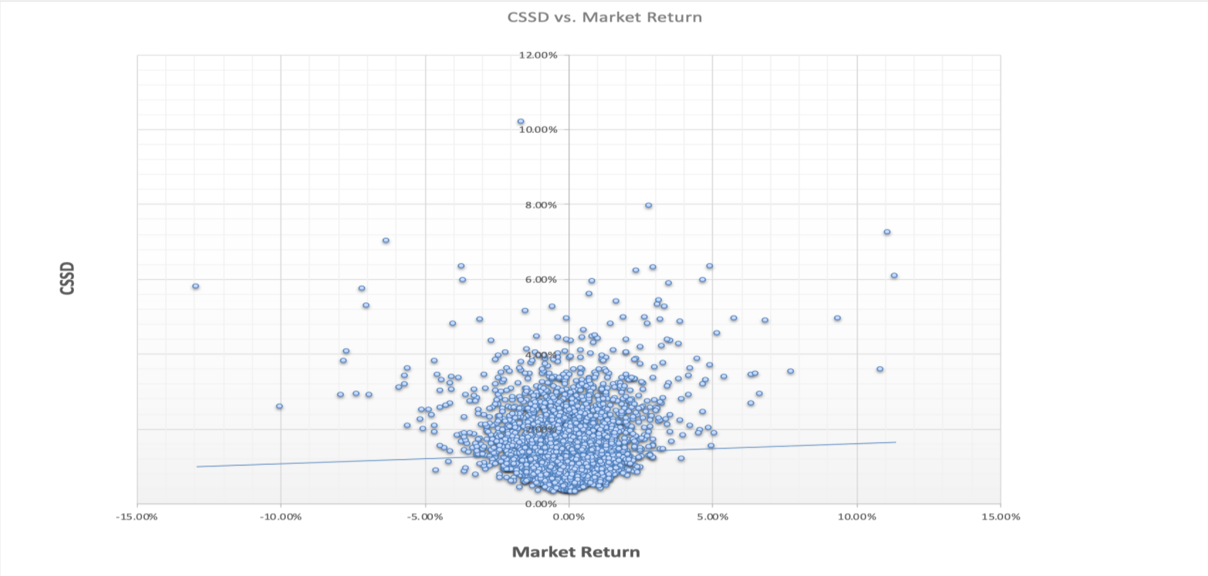
Cheng et al. (2000) have proposed a slight change in equation (4):

$$CSAD_t = y_0 + y_1 R_{m,t} + y_2 R^2_{m,t} + \varepsilon_t$$

in order to capture the herd behavior in an up market. The importance of this regression model is that the market return does not need to be in the upper or lower tail of extreme market movements to pick up on herd behavior within the market.

A third regression introduced by Chiang and Zheng (2010) shows that the first inclusion of $R_{m,t}$ as a regressor. By adding these additional regressors we are able to strengthen the explanatory power of the regression model. This regression specifically will test for asymmetry between the two variables, the market return and dispersion values. They propose two separate ways to test or herding behavior, and suggest splitting the data into two categories; the market behavior on up days and the market behavior on down days. The original equation is as follows:

Figure 1. Relationship between (CSSD) , and the market return



$CSAD_t = y_0 + y_1 R_{m,t} + y_2|R_{m,t}| + y_3 R^2_{m,t} + \varepsilon_t$

They have proposed these changes in order to capture the market on down-days and-up days and thus replace $R_{m,t}$ with $(1\text{-}D).R_{m,t}$ and $D.R_{m,t}$ and to replace $R^2_{m,t}$ with $(1\text{-}D) R^2_{m,t}$ and D $R^2_{m,t.,}$ where the dummy variable, D, has a value of 1 when $R_{m,t} < 0$, and a value of 0 otherwise.

RESULTS AND ANALYSIS

To begin our analysis, we have chosen to present visual representations of the two return dispersion measures that have been calculated in this study. The first visual graph is of the cross-sectional standard deviation (CSSD), and this is the dependent variable of the graph and plotted on the y-axis. The independent variable is the market return and it is plotted along the x-axis. The graph does not show a strong relationship that fits the trend line of the data. Many of the data points exceed 2%, which is a large deviation in terms of financial returns. Another noticeable characteristic of this dataset is the number of outliers. Visually, one can see that there are frequent instances where the individual stock returns deviated upwards of 6% and even deviated as far as 10% away from the market return.

In the second visual representation we have replaced the cross-sectional standard deviation (CSSD) with the cross-sectional absolute deviation (CSAD).

Figure 2. Relationship between (CSAD) , and the market return

It is plotted on the same axis as the first graph. When looking at the graphical representation of the cross-sectional absolute deviation, one can see very similar results. The deviation of individual stock returns continues to show weak signals of herd behavior in the Dow Jones Industrial Average. Each graph does have a cluster of data point around the center of the graph, leading one to believe that there is a slight herding relationship between the two variables — market return and individual stock returns. Later, in the regression analysis sections, it will be shown that the R-squared values do account for some of the data points in this study, but we conclude that they are not statistically significant.

Summary Statistics

Table 1 provides summary statistics of the CSSD, CSAD, and the market return over the entire twenty-year time period, with 5282 data points. The mean for the market return over the entire twenty years period is 0.03%. As mentioned earlier, at the start of the dataset the market had just come off the tech bubble popping, and shortly after experienced the Global Financial Crisis. Following the crisis, the market had experienced the longest Bull Run in the history in itself.

One can assume that with having nearly half of the period was in a bull market, the mean would have higher value; however, COVID-19 Pandemic caused a large correction in the DJIA in the final year of the period, which significantly affects the mean over time. Furthermore, by reviewing the basic statistical summary in Table 1, one can see that the mean for both CSSD and CSAD are, 1.34% and 0.93%, respectively. The mean of each of these measurements is significantly larger than the mean of the market return. Each of these values lie within, or extremely close to one standard deviation of the market return mean, however, one standard deviation of the market value is 1.21% which is a significant percentage difference in return value.

Table 1. Summary Statistics of CSSD & CSAD				
	Mean	Std. Dev.	Minimum	Maximum
$R_{m,t}$	0.03%	1.21%	-12.93%	11.37%
CSSD	1.34%	0.78%	0.31%	10.21%
CSAD	0.93%	0.53%	0.23%	6.42%

Dummy Variable Regression

We have also used CSSD and CSAD in various regression models to further investigate the herding behavior within the market. In the first regression, we have used two dummy variables describing whether the market return fell within the lower tail or the upper tail of the return distribution. This regression has attempted to capture the herding tendencies of investors during extreme market movements. Earlier in the study we defined the upper tail as any return that is 1.7% and higher, and we defined the lower tail as any return that was -1.83% and lower (Dang and Lin 2016). Table 2 shows the results of the regression analysis. The coefficient estimates of the left tail and right tail are positive, thus implying that per change in the market return there is not a drastic or extreme response in the asset prices of the securities that comprise the index; and furthermore, they imply that if there is any change. The value of R-squared, 0.1394, indicates that the regression analysis is accounting for around 14% of all the variations in data

Table 2: Dummy Variable Regression			
	Left	Right	Intercept
Coefficient	0.007751083	0.011381754	0.012437045
Standard Error	0.000460344	0.00045952	0.0001056
R-Square	0.139443911	0.007280243	#N/A
F-Stat.	427.7027475	5279	#N/A
	0.045338154	0.279797261	#N/A
T-Stat.	16.83760502	24.76879225	
P-value	0.00%	0.00%	

points. Furthermore, the p-values are extremely low and near zero, indicating the significance of these values. In the context of study by Christie and Huang (1995), the DJIA behaved rationally and there are no signs of herding over this twenty-year period. The dispersion values continually increased in the left tail and the right tail. This implies that the investors do not herd toward the market signals, and ignore the opinions of key influencers.

CSAD Regression Analysis

We have further expanded the regression analysis by utilizing the cross-sectional absolute deviation (CSAD), and added a third regressor. This time, instead of using dummy variables to determine when the market was within extreme market movements, we were able to capture whether or not there is any non-linearity between the two variables. We shifted our focus towards the return measures and regress the absolute market return and squared market return against the cross-sectional absolute deviation return values.

The regression estimates are shown in Table 3. There is a negative coefficient for the return squared component, with a value of -0.0149. The presence of a negative estimate for the quadratic term (y_2) indicates that there is a potential that herding behavior is present in the market. However, the R-squared value is not that significant, as it is only explaining around 25% of variations in CSAD. The p-value is relatively high which indicates the estimated coefficient is not significant. There is a stronger potential of herding in the market when considering the CSAD model due to the non-linear relationship present in the data set. The model specifically tested for a nonlinear relationship, which would indicate that as the market return increased (decreased) the dispersion measurement would not move in comparison with the market. It instead would move in an opposite direction, increasing the dispersion value. Therefore, the presence of a non-linear relationship is a strong indication of herd behavior.

Asymmetry Regression

Table 3: CSAD Regression Table			
Components:	$(\text{Return})^2$	ABS (return)	Intercept
Coefficient	-0.01497852	0.293223448	0.007041245
Standard Error	0.20011615	0.012207646	9.72089E-05
R-Square	0.253024616	0.00460304	#N/A
F-Stat.	894.0836443	5279	#N/A
	0.037887656	0.111851355	#N/A
T-Stat.	-0.074849134	24.01965595	72.43417061
P-value	94.03%	0.00%	0.00%

The third regression that has been conducted is the most powerful of the three in regards to explanatory power. There are two versions of the last regression. The first version uses both the market return and absolute market return, as well as the quadratic function from the previous regression model. This model has the potential to capture the market behavior during any conditions, both up and down. More specifically, for every change in the market return, CSAD will change by $Y_2 + Y_1$ for any value that $R_{m,t}$ is positive and change by $Y_2 - Y_1$. Moreover, this means that asymmetry can be quantified as the ratio of $\frac{Y_2+Y_1}{Y_2-Y_1}$ (Dang and Lin 2016). The estimates of this regression are presented in table 4.

The coefficient for the quadratic term is negative, which in turn does present a potential for herding behavior to be present in the market. The R-squared value is still not entirely significant with a value of about 25% of the data points in the study. Although the R-squared value is not significant the p-value does indicate there is some significance to the coefficient estimate for the regression.

The second version of this regression model divides the performance on the market into two separate groups; days when the market is up and days when the market is down. By using dummy variables to create a distinction between the two conditions. The estimates for the regression are presented in table 5. The coefficient estimates for this regression are mainly positive with only one of the regressors having a negative estimate. The dummy term of the regression has the highest p-value levels showing the highest significance, however, the R-squared values range from 2% to 27% of the data points within the dataset. The significance level given by the p-values for the quadratic term and the dummy variable give an indication that there is not a strong significance to these coefficients.

Table 4: Asymmetry Regression Table

Equation 5	RET^2	ABS RET	RET	Intercept
Coefficient	-0.031451987	0.295111599	0.02442784	0.007022484
Standard Error	0.199756386	0.012190531	0.005244672	9.71025E-05
R-Square	0.256082274	0.004594045	#N/A	#N/A
F-Stat.	605.6235316	5278	#N/A	#N/A
	0.038345506	0.111393505	#N/A	#N/A
T-Stat	-0.157451725	24.20826442		
P-value	87.49%	0.00%		

Table 5: Revised Asymmetry Regression (Up & Down)

	D	D.R^2m,t	(1-D).R^2m,t	D.Rm,t	(1-D).Rm,t	ABS RET	Intercept
Coefficient	0.2552528	-0.3032706	0	0.1014482	0.24	0.00	0.01
Standard Error	0.2865408	0.2782435	0	0.0243679	0.02	0.00	0.00
R-Square	0.2574049	0.0045908	#N/A	#N/A	#N/A	#N/A	#N/A
F	365.76286	5276	#N/A	#N/A	#N/A	#N/A	#N/A
	0.0385436	0.1111955	#N/A	#N/A	#N/A	#N/A	#N/A
T-Stat	0.8908078	-1.089947	#DIV/0!	4.163191598	14.13927441	3.059145	
P-value	37.31%	27.58%	#DIV/0!	0.00%	0.00%	#N/A	

CONCLUSION

After exploring the scope of the relationship between the market returns and return dispersion, we can conclude that there is not enough evidence to support the statement that herding is prevalent in the Dow Jones Industrial Average between the years of 2000 and 2020. There was not sufficient evidence presented by the Christie and Huang (1995) cross-sectional standard deviation dummy regression calculated in this study to prove that herding is present in the market during extreme market movements. The only regression that presents a case for herding being present in the market index is the first CSAD regression performed in this study. Although, the R-squared value is not nearly as significant, the coefficient estimate of the quadratic term is still negative and this represents a non-linear relationship between CSAD and the market return. The asymmetric regression is quite similar to the non-linear test in results. The quadratic term does have a negative estimate. However, the R-squared value is not strong enough to indicate that the results present a potential for herd behavior in the market place. The second version of the asymmetric regression does not provide any sufficient evidence that herding occurs on down or up days in any specific pattern.

Based on the findings in this study it is reasonable to believe that the market behaves in a rationale sense, ignoring the signals of the market and its' largest influencers. In other words, the Efficient Market Hypothesis stands. In opposition to the literature for behavioral finance, individual investors are able to, based on the empirical evidence, look past their natural human biases and heuristics. This is not to say that spurious herding does not still occur in today's markets, but there is no evidence that intentional herding is a consistent phenomenon that occurs in the market.

When we examine the meaning of the coefficients throughout each of the regression examples, we noticed that only two of them points towards potential herd behavior in the market.

As referenced earlier in the essay, the CSAD regression has the strongest indication of herding in the market, due to the negative coefficient. This means that for any move in the market return that is positive, the cross-sectional absolute deviation will move in the opposite direction closing the spread between the market return and the stock returns. The second most indicative regression was the asymmetric regression, which also presents a negative coefficient leaning towards the presence of potential herd behavior in the market. The p-values do show a strong significance; however, the R-squared values show that the regressions don't represent enough of the dataset to mean too much. The first regression points to a positive relationship between the market return and the dispersion of stock returns, meaning that as the market return increases, so does the deviation of other stock returns from the market return. For example, during the days of market down, the individual stock returns are likely to move in a positive direction, and on market up days the individual stock returns are likely to move in a negative direction. The CSSD regression has a positive coefficient, which indicates that the market return and the dispersion value move in relation with each other. In other words as the market price increases so does the dispersion of the individual stock returns. The fact that the R-squared values don't give sufficient backing to the indication of herd behavior in the Dow Jones Industrial Average it leads us to believe the market behaves rationally.

In regards to the earlier questions asked in this study, the return dispersion of individual assets does not present itself as a strong indicator of herding in the market. Instead, it serves as proof that investors do not act on market signals, and instead develop their own investment decision process with the information at hand. As for the second question, there is not sufficient evidence of herding being present in the market; so, it is not reasonable to assume there is or is not an effect on overall asset prices. Instead, the evidence presented in this study lead to the conclusion that the market is overall rational. The investors, individually, use the readily available public knowledge to forge their own investment thesis. The effect of financial information is priced into the market; however, that does not always mean spurious herding cannot occur in this situation. Volatility will always be present in the market, but that does not always mean it is directly correlated to herd behavior.

REFERENCES

Bikhchandani, S., & Sharma, S. (2001). Herd Behavior in Financial Markets. *IMF Staff Papers*, 47, 279-310.

Chang, Eric C., et al. "An Examination of Herd Behavior in Equity Markets: An International Perspective." *Journal of Banking & Finance*, vol. 24, no. 10, Oct. 2000, pp. 1651–1679.

Christie, W. G., & Huang, R. D. (1995). Following the Pied Piper: Do Individuals Returns Herd around the Market? *Financial Analysts Journal*, *51*, 31-37.

Dang, Ha V., and Mi Lin. "Herd Mentality in the Stock Market: On the Role of Idiosyncratic Participants with Heterogeneous Information." *International Review of Financial Analysis*, vol. 48, 2016, pp. 247–260., doi:10.1016/j.irfa.2016.10.005.

Dhall, Rosy, and Bhanwar Singh. "The COVID-19 Pandemic and Herding Behaviour: Evidence from India's Stock Market." *Millennial Asia*, vol. 11, no. 3, 2020, pp. 366–390., doi:10.1177/0976399620964635.

Fama, E. F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, *25*, 383-417.

Gębka, Bartosz, and Mark E. Wohar (2013),"International Herding: Does It Differ across Sectors?" Journal of International Financial Markets, Institutions & Money, vol. 23, Feb. 2013, pp. 55–84

Igual, Manuel G., and Teresa C. SantaMaria. (2017), "Overconfidence, Loss Aversion and Irrational Investor Behavior: A Conceptual Map." International Journal of Economic Perspectives, vol. 11, no. 1, Mar. 2017, pp. 273–290Qawi, Raluca Bighiu (2010). "Behavioral Finance: Is Investor Psyche Driving Market Performance?" IUP Journal of Behavioral Finance, vol. 7, no. 4, Dec. 2010, pp. 7–19.

DECISION SCIENCES INSTITUTE

Quantum Computing-inspired Intelligent Learning for Data-driven Combinatorial Optimal Decision Making in Operations Engineering

Pan Zou
Georgia Institute of Technology
Email: zoupzp123@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

Operations Engineering domain is mainly concerned with the application of advanced analytical methods to help improve the decision-making process in operational problems under complex and most of time dynamic environment. This paper reviews and analyzes the existing gaps and challenges, and further proposes a significant trend and framework of applying quantum computing inspired, context-aware intelligent learning in the data-driven decision-making paradigm to enhance the operation engineering problem solving.

KEYWORDS: Quantum Computing, Intelligent Learning, Data-driven Decision Making, Combinatorial Optimization, and Operations Engineering

BACKGROUND & MOTIVATION

Industries nowadays, along with their operation environments, are confronting more and more disruptive changes. These changes are not only caused by the up-and-coming automation and communication technology and the corresponding massive production data, but also by the external limitation and adjustment due to the global public health emergency COVID-19 pandemic and its associated long-term impacts. Combinatorial optimal decisions are widely observed in operations engineering applications, ranging from modular design (Borissova and Keremedchiev, 2019) to operations planning and scheduling (Gao et al., 2017), and to logistics and supply chain engineering (Vogiatzis and Pardalos, 2013). These decisions essentially entail typical combinatorial optimization problems. Critical research problems that are imperative regarding the complexities of combinatorial optimal decisions in operations engineering problem solving are observed as the following:

(1) Computational Efficiency

Combinatorial optimization is proven to be an NP-hard problem, for which computational efficiency is recognized as a classical challenge. It is associated with the well-known combinatorial explosion problem due to the inherent enumeration nature of combinatorial optimization (Mukelabai et al., 2018). Due to the increasing complexities and dynamics in the operation environments, it is one of the forever pursuits to improve the computation efficiency in operations problem solving. Conventional mathematical programming approaches to combinatorial optimization problem formulation and model solution include exact methods, approximation methods and heuristic-based methods (Fernandes Muritiba, 2010). These combinatorial optimal decision models and solutions are significantly limited by the growing problem size and complexity incurred in practical applications (Neumann and Witt, 2010).

(2) Domain Problem Context Awareness

The conventional wisdom of combinatorial optimization problem solving is consistent with the basic modeling framework of mathematical programming, which is based on the appropriate simplification and approximation to abstract domain problems. Meanwhile, sufficient practical cases have revealed that many types of operational optimization problems are solved better by leveraging their particular domain problem contexts and decision structures (Jouglet and Carlier, 2011). Some algorithms, though designed to be general, are found, in hindsight, to be empirically more efficient for certain types of problems (Blum et al., 2011). These solvers, however, are often based on rules and special heuristic design, which highly relies on expert knowledge and involves preference bias. As a result, the computing is often iterative and serialized, suffering limitations in scalability, adaptation ability, converging speed, and accuracy. Recent years have witnessed a rapid expansion of the machine learning frontier for solving these combinatorial optimization problems (Mazyavkina et al., 2020), especially when large amount of data and distributions of problem instances are available. Given its high flexibility, approximation nature, and self-learning paradigm, machine learning is particularly attractive to the call for enhancing domain problem context awareness in combinatorial optimization modeling. It is thus important to examine the potential of machine learning in learning universal high-quality heuristics for discovering new advanced algorithms, especially for applied operational problems (Bengio et al., 2021).

(3) Data-driven Optimal Decision Making

Practical operational problems require consideration and balancing of complex and dynamic operational conditions and situations (Ermon et al., 2012). Optimal operational decision making requires not only the optimization capability based on a rigid formulation of the problem domain, but also the self-adaptability to proactively respond to the dynamic nature of the operational environment (Cruz et al., 2011; Artikis et al., 2014; Sengupta et al., 2017). With the pervasive connectivity and extensive smart sensing technologies deployed in today’s operations systems, massive operational data and user generated contents can be collected uninterruptedly (Yang et al., 2019). Recently data-driven decision making has attracted much attention as an emerging paradigm to make better-informed decisions (Jin et al., 2019). By exploiting huge, versatile, and highly contextualized operational data, operation engineers can harness their organization’s competitive edge by uncovering patterns, novel insights, and domain knowledge through data-driven decision making (Zhao et al., 2007). Nonetheless, there is always sizeable noise and preference bias contained in the massive data associated with complex domain problem contexts. It is of practical significance to examine the essential impacts and the best form of data-driven operational decision making to embrace the benefits of the emerging data analytics paradigm.

This paper aims to propose a significant trend of applying quantum computing inspired, context-aware intelligent learning in the data-driven decision-making framework, for the sake of enhancing the operation engineering problem solving. The rest of the paper has the following sections: the next “literature review” section will talk about the related theoretical foundation and technical rationality of this framework. The high-level framework structure will be addressed in the “Methodology” section, while some possible application instance is given in “Application Example” section. In the end, the “Discussion and Conclusion” section will summary some potential challenges and possible future work.

LITERATURE REVIEW

Quantum computing has been foreseen as the next-generation computation paradigm, as it offers a fundamentally different type of solutions to computational problems and thus to bring a significant boost in computation efficiency (Duan and Raussendorf, 2005). Nonetheless, there is still a long way to go for it to be practically ready on quantum computers due to limitations of hardware resources and physical infrastructures (Fellous-Asiani et al., 2020). But rather, quantum-inspired techniques are suggested to be more practically feasible and promising, that is, new quantum-inspired computational methods can be originated from the basic principles of quantum mechanics to enhance classical algorithms (Arrazola et al., 2019). To address the foundation of technical approaches in this proposed framework, the paradigm of quantum computing and the practical practices using quantum-inspired evolutionary algorithms are reviewed below.

Quantum Computing

The idea of quantum computation is that if the particles that follow their natural quantum-mechanical behavior can be directly regarded as a "quantum computer", then this "quantum computer" appears to simulate the computation of a quantum system exponentially more efficiently than performing it with a classical computer. Therefore, by sophisticatedly utilizing the multi-particle systems, their natural quantum behaviors can enable other computations to be exponentially more efficient. In theoretical quantum computation modeling, Q-bit (quantum bit) is first as least as powerful as a classical bit, as it could be in any Boolean state '0' or '1', and further, it can be in a linear combination of the basis states $|0\rangle$ and $|1\rangle$, also known as quantum superposition. A general form of single Q-bit quantum state $|\psi\rangle$ defined in Hilbert space is:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \alpha\begin{bmatrix}1\\0\end{bmatrix} + \beta\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}\alpha\\\beta\end{bmatrix} \qquad (1)$$

where the probability amplitudes $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2$ gives the probability of the Q-bit in the state $|0\rangle$, and $|\beta|^2$ the state $|1\rangle$, which naturally leads to $|\alpha|^2 + |\beta|^2 = 1$. The combined state of two or more Q-bits is the tensor product of them, for instance, the vector representation of two Q-bits:

$$|\psi_1\psi_2\rangle = |\psi_1\rangle \otimes |\psi_2\rangle = \begin{bmatrix}\alpha_1\\\beta_1\end{bmatrix} \otimes \begin{bmatrix}\alpha_2\\\beta_2\end{bmatrix} = \begin{bmatrix}\alpha_1\alpha_2\\\alpha_1\beta_2\\\beta_1\alpha_2\\\beta_1\beta_2\end{bmatrix} = \alpha_1\alpha_2|00\rangle + \alpha_1\beta_2|01\rangle + \beta_1\alpha_2|10\rangle + \beta_1\beta_2|11\rangle \qquad (2)$$

Quantum computation implements computation via manipulating quantum states by using quantum circuits that consists of registers, ancillas, and quantum gates. Some of the most widely used ones can be presented in the form of unitary matrices:

$$Pauli\ Matrices: X = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}; Y = \begin{bmatrix}0 & -i\\i & 0\end{bmatrix}; Z = \begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}. \qquad (3)$$

$$Hadamard\ Gate: H = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix} \qquad (4)$$

$$Controlled\ NOT\ Gate: CNOT = CX = \begin{bmatrix}1 & 0 & 0 & 0\\0 & 0 & 0 & 0\\0 & 0 & 0 & 1\\0 & 0 & 1 & 0\end{bmatrix} \qquad (5)$$

Two most spectacular examples of carefully designed quantum system following laws of quantum mechanics are Shor's factoring algorithm (Shor, 1994) and Grover's database search algorithm (Grover, 1996). Their studies show that quantum computers are in some sense more powerful than classical computers at least with respect to solving some specific problems. However, the practical execution of quantum computer is still in its very early infancy (Weber et al., 2010; Zhang, 2011).

Figure 1. Quantum Circuit of Shor's Factoring Algorithm Implementation

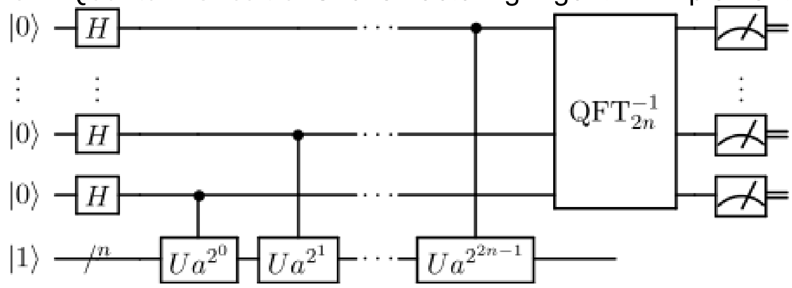
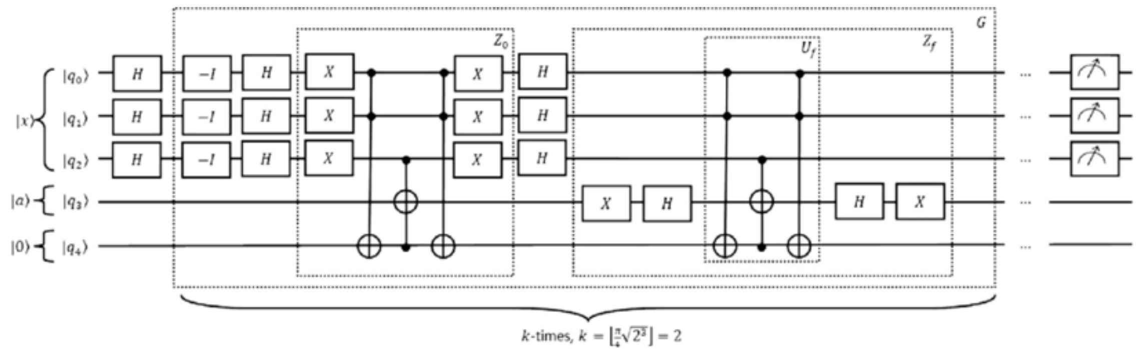


Figure 2. Quantum Circuit of Grover's Search Algorithm Implementation



Quantum-inspired Evolutionary Algorithms

Quantum-inspired computing has emerged as a novel computational paradigm and is characterized by applying computational method inspired by principles and concepts of quantum mechanics while using classical algorithms to evaluate candidate solutions (Moore and Narayanan, 1995). It has shown great potential on the complexity class NP (the class of exponential search problems), whereas not requiring to be implemented on an actual quantum computer. For example, inheriting most of the features of evolutionary algorithms, quantum-inspired evolutionary algorithms introduce quantum concepts and principles in the computation, such as qubit, superposition, parallelism, and quantum gates, yet only require a classical computer rather than quantum mechanical hardware. In a typical quantum-inspired evolutionary algorithms (QIEAs) (Han and Kim, 2002), a Q-bit based individual is often represented as a string of *n* Q-bits: $\begin{bmatrix}\alpha_1 \\ \beta_1\end{bmatrix}\begin{matrix}\alpha_2 \\ \beta_2\end{matrix}\begin{matrix}\cdots \\ \cdots\end{matrix}\begin{bmatrix}\alpha_n \\ \beta_n\end{bmatrix}$, where the probability amplitude $\alpha_i, \beta_i \in \mathbb{C}, i \in \{1,2,\dots,n\}$ also conforms to $|\alpha_i|^2 + |\beta_i|^2 = 1$, while $|\alpha_i|^2$ and $|\beta_i|^2$give the respective probability of the i^{th} Q-bit in the state $|0\rangle$ and the state $|1\rangle$.

During the problem-solving procedure, the fitness evaluation procedure requires the "Measurement" of quantum state first, but with the assumption of not causing "wave function collapse" which would "destroy" the quantum system in quantum physics and thus the narrow definition of quantum computing. The crossover and mutation procedures of the population are implemented by quantum gates, for instance, quantum rotation gate $Rot(\Delta\theta_i)$. The update of a single Q-bit aims to adjust the probability of the state in one specific basis state based on the evaluation result, and it follows:

$$Rot(\Delta\theta_i) = \begin{bmatrix} \cos(\Delta\theta_i) & -\sin(\Delta\theta_i) \\ \sin(\Delta\theta_i) & \cos(\Delta\theta_i) \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} \alpha_i' \\ \beta_i' \end{bmatrix} = \begin{bmatrix} \cos(\Delta\theta_i) & -\sin(\Delta\theta_i) \\ \sin(\Delta\theta_i) & \cos(\Delta\theta_i) \end{bmatrix} \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \quad (7)$$

The absolute value of $\Delta\theta_i$ (a definable parameter that is associated with the converging speed) can be 0 or 0.01π, and the lookup table helps decide the value and sign of $\Delta\theta_i$ based on the quadrant the Q-bit located. Most of the QIEAs have been tested on Knapsack problems, which is one well-known combinatorial optimization problem, and have shown good performance in problem solving. The advantages of quantum-inspired methods are widely reported over classic computational algorithms in terms of exponential improvement of computational efficiency (Xiong et al., 2018).

Figure 3. Pseudo Code of QIEA

```
Procedure QIEA
BEGIN
t ← 0
i)   Initialize Q(t)
ii)  Make P(t) by observing the states of Q(t)
iii) Evaluate P(t)
iv)  Store the best solutions among P(t) into B(t)
     WHILE (!termination-condition) DO
     BEGIN
     t ← t + 1
     v) Make P(t) by observing the states of Q(t-1)
     vi) Evaluate P(t)
     vii) Update Q(t) using Q-gates
     viii) Store the best solutions among B(t-1) and P(t)
     into B(t)
     ix) Store the best solution b among B(t)
     x) IF (migration-condition)
        THEN migrate b or b_j^t to B(t) globally or locally,
     respectively
     END
END
```

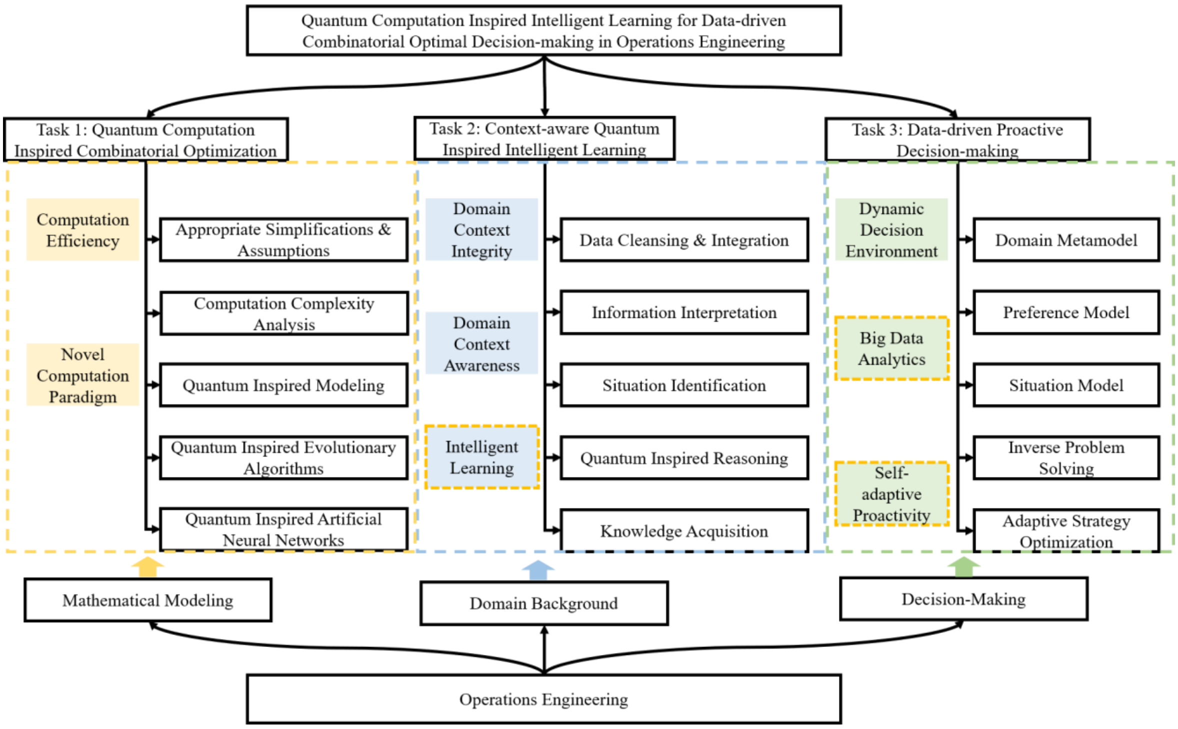
```
Procedure Make(x)
BEGIN
i ← 0
     WHILE (i < m) DO
     BEGIN
     i ← i + 1
     IF random[0,1] < |β_i|^2
     THEN x_i ← 1
     ELSE x_i ← 0
     END
END
```

```
Procedure Update (q)
BEGIN
i ← 0
     WHILE (i < m) DO
     BEGIN
     i ← i + 1
     Determine Δθ_i with the lookup table obtain
     (α'_i, β'_i) from the following:
          IF (q is located in the first/third quadrant)
          THEN [α'_i, β'_i]^T = U(Δθ_i)[α_i, β_i]^T
          ELSE [α'_i, β'_i]^T = U(−Δθ_i)[α_i, β_i]^T
     END
     q ← q'
END
```

METHODOLOGY

The research methodology here is to create mathematical & computational models and frameworks to advance the basic research for solving the current technical challenges in operation engineering with particular focus on the efficiency improvement using quantum computation inspired combinatorial optimization, the completeness of context-awareness via intelligent learning, and the self-adaptivity of decision-making process driven by the massive contextual data collected from the operational environment, as shown in Figure 4. Both qualitative and quantitative methods would be employed, along with the standard benchmark problems testing and actual industrial case studies to test research hypothesis, validate theoretical models and algorithms, and assess technical concepts.

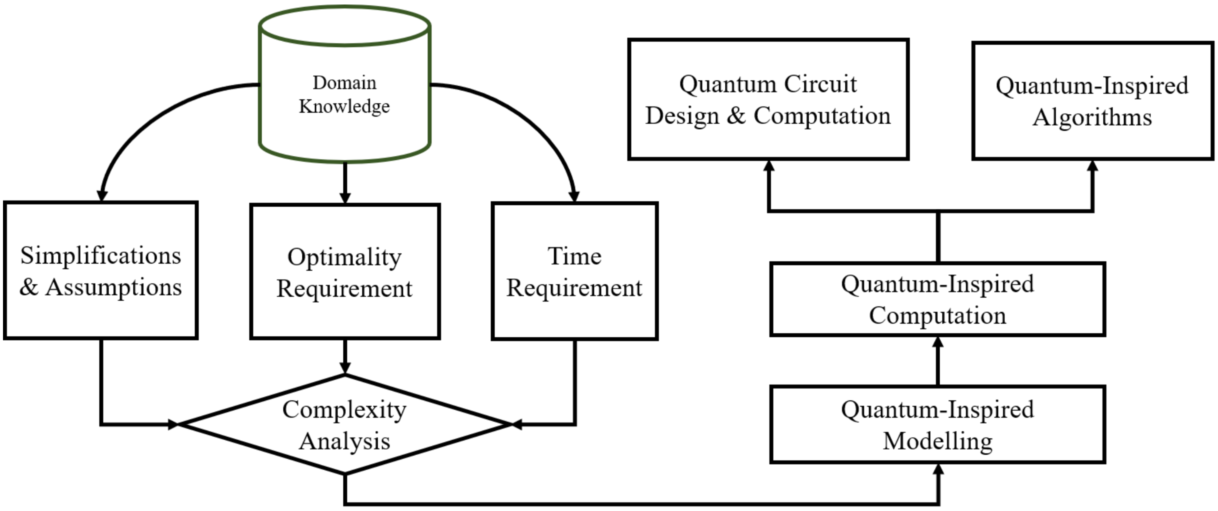
Figure 4. A Research Roadmap of the Proposed Framework



Quantum Computing-inspired Combinatorial Optimization

The core idea of quantum computing is to carefully engineer quantum circuit systems that consist of multiple particles and leverage their natural quantum behaviors and quantum parallelism, to achieve exponentially more efficient computation (Williams, 2010). Quantum-inspired modeling mimics the quantum characteristics, such as quantum bit (namely, qubit or Q-bit), quantum parallelism and quantum gates (Q-gate). Inheriting both the features from quantum-inspired computing and evolutionary algorithm, quantum-inspired evolutionary algorithms have recently attracted much attention (Zhang, 2011). By introducing Q-bit to the individual representation, these heuristic algorithms bring exponentially higher diversity to the entire population (Han and Kim, 2002), and therefore broaden the searching horizon (Han and Kim, 2000). Quantum-inspired heuristic models have shown very competitive performance on several simple yet typical combinatorial optimization problems comparing to conventional computation methods (Han et al., 2001; Kim et al, 2006). Additionally, these heuristic algorithms are often easy to implement and can be tailor-made for complex domain problems including operations engineering (Patvardhan et al., 2015).

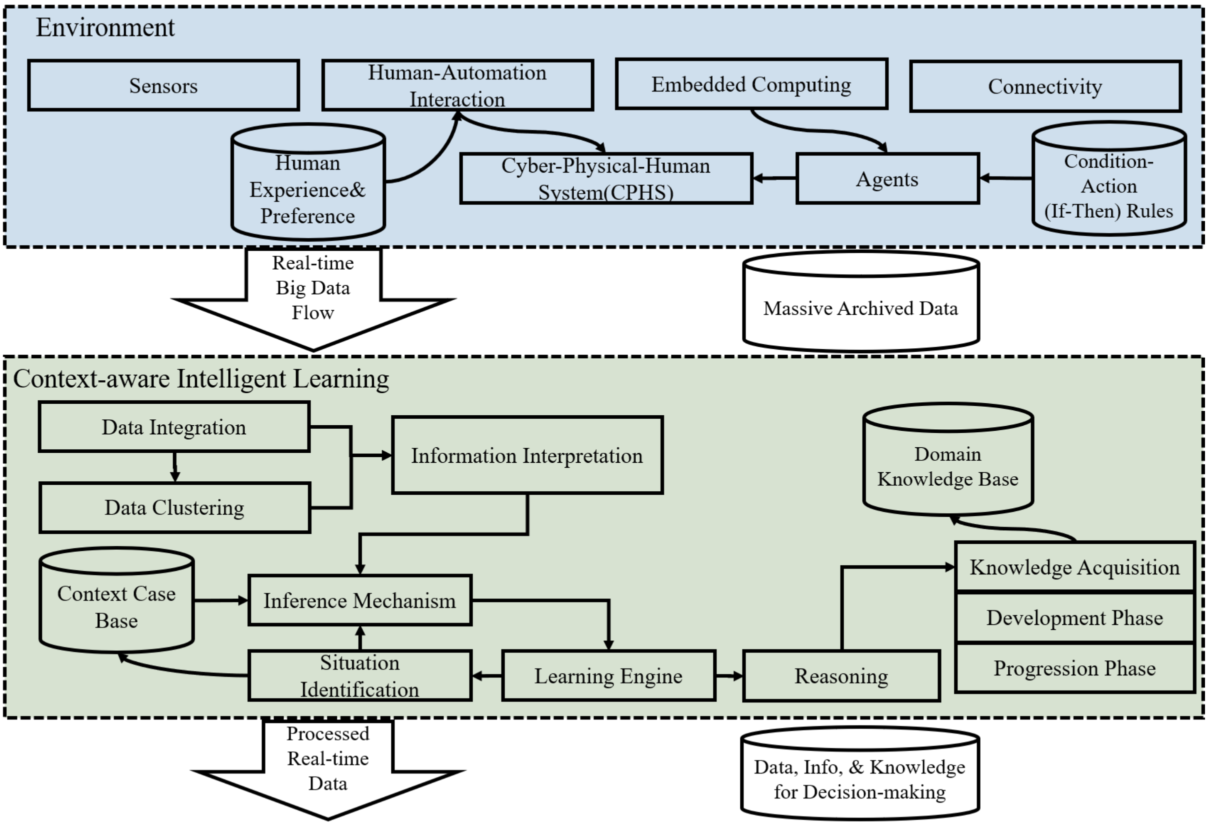
Figure 5. Modelling of Quantum Computing-inspired Combinatorial Optimization



Context-aware Quantum-inspired Intelligent Learning

Sensing and communication technologies are being used in daily operational activities. In the paradigm of "Internet of Things" (IoT), these technologies are monitoring the real-world environment, the digital processing and connecting between multiple agents in the system, as well as the human-automation interactions, which can be influenced by domain experience and personal preference. The comprehensive awareness of the current context needs to be learned more intelligently, which requires data processing, information interpretation, situation identification, reasoning, as well as knowledge acquisition. Once the data is extracted from the sensing source, data processing, which includes data cleansing, data transformation, data clustering, etc., needs to be done to autocorrect the explicit errors and noise, filter the potentially useful information out, and thus reduce the complexity of posterior computation. The capabilities of situation identification and inductive reasoning would be facilitated by the learning engine that learns from the massive real-time monitoring data and archived historical data. Since majority of the monitoring data is time dependent, the integrity analysis of the current context awareness will be processed to evaluate the sufficiency of data and decide optimal learning intervals. In the learning engine, the task of situation identification can be formulated as a mixed multi-input classification as well as anomaly detection problem, since the type of data being collected can cover both categorical (discrete) and continuous ones and a well-generalized classification model will predict outliers the same way it predicts the class label. Once the satisfactory criteria have been met, the output of the quantum-inspired neural network model could serve the role of predictor in the reasoning model and enhance the comprehensive performance of the reasoning section. All the processed data, information, and validated knowledge after the context-aware intelligent learning system would be structured, stored, and transmitted for next-stage decision-making process.

Figure 6. Intelligent Learning Engine



Data-driven Proactive Decision Making

Data collection and analysis have played an increasingly prominent role in operation engineering decision-making process, together with the proliferating complexities for building decision models, due to the rapid changing operational environment and external environment. Instead of building the conventional model-driven decision-making with exponentially growing complexities, data-driven decision-making system with self-adaptability would be a more rigorous solution. As the large amount of added value information enhance context awareness, meaningful insights about potential future problems could be obtained for a proactive manner of decision-making, especially for those human-in-the-loop scenarios. Additionally, the predictive simulation required by the proactivity can be escalated further within the quantum-inspired computation paradigm. Any other underlying sub-optimization problems here that can be modelled into ones that belong to the paradigm of quantum inspired computation would be handled correspondingly to maximize the computation efficiency.

APPLICATION EXAMPLE

The abstraction from real-world operational problem to generic combinatorial optimization problem requires appropriate simplifications and assumptions as well as domain knowledge and modelling experience. If the problem happens to fall in any specialty of quantum computing, appropriate quantum-inspired modelling methods and the corresponding quantum-inspired computing algorithm should be further justified and weighed in the next step. Here, an application of quantum-inspired combinatorial optimization in modular design is given as a possible practice example.

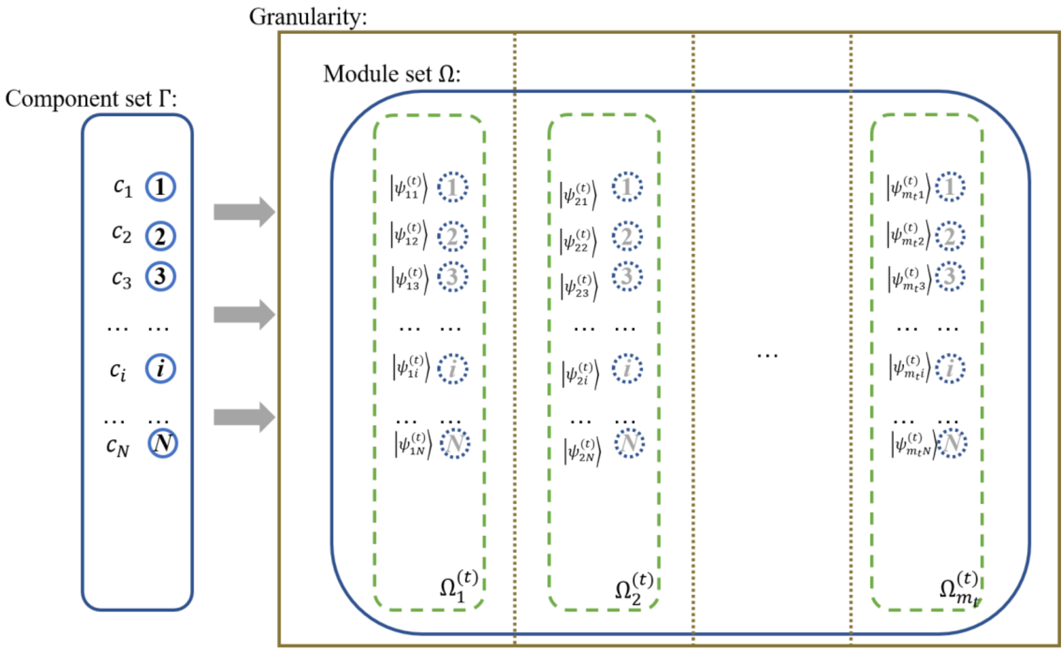
Modular Design

The main idea in modular design is to identify and group small product components into several independent modules with the aim of minimizing between-module interactions and maximizing within-module interactions. Modular design is classified as a typical application domain of combinatorial optimization problems, which can be defined as follows: Given a group of N product components Γ and a fitness function f to evaluate the overall optimality of modules, the goal is to find the best clustering strategy to maximize the fitness. A feasible strategy includes the determination of the number of clusters M and the assignment of the objects correspondingly. Theoretically, M can be any integer value between 1 and N.

$$
\begin{gathered}
max\ (f(\Omega_1, \Omega_2, \dots, \Omega_M)) \\
subject\ to\colon \bigcup_{k=1}^{m} \Omega_k = \Gamma \\
\bigcap_{k=1}^{m} \Omega_k = \emptyset \\
\forall\, k \in \{1, 2, \dots, M\}, \Omega_k \neq \emptyset \\
M \in \{1, 2, \dots, N\}, M \in \mathbb{Z}
\end{gathered}
$$

The quantum-inspired modelling of this problem is to use the concept of quantum superposition for the state representation.

Figure 7. Quantum-inspired Modeling for Modular Design Problem



DISCUSSION AND CONCLUSIONS

Due to the ever-increasing consumer demands and the need for industries to be efficient, future data will be packed with uncertainties. These uncertainties will make the process of decision making extremely challenging for current decision makers as the decision space will increase exponentially and they may not be unable to arrive at a feasible solution with practical time. Furthermore, conventual optimization techniques will also be restricted in their ability to cope with these complexities due to the inefficient solution space exploration process. The proposed quantum computation inspired intelligent learning for data-driven decision-making approach goes beyond the conventional decision-making approaches in the following aspects:

(1) Quantum and quantum-inspired computing can bring a very promising computation efficiency enhancement. The usage of quantum inspired operators and quantum measurement operators enables more efficient exploration of the search-space minimizing the risk of exploration-exploitation trade-off normally encountered in classical evolutionary algorithms.
(2) By enabling the learning engine to self-extract context, we will alleviate the sole reliance upon human engineers and operators for time-sensitive decision making. Furthermore, the knowledge acquisition module will be developed based on precise historical data and human expertise and will be capable of extracting relevant information from new data and update the exploration-exploitation trade-off strategies.
(3) Unlike conventional model-driven decision-making process, the proposed methodology can make optimal decision when provided with the data and information that narrates the dynamics and uncertainties of the operation environment, with the support of quantum-inspired prediction modelling. An advantage of the proposed methodology is its ability to evolve new strategies using predictive simulations and hypothesis testing when accumulated dynamics and uncertainties derail the effectiveness of current decision strategies. The proposed quantum computation inspired intelligent learning for decision making framework is a novel approach for enabling flexibility and robustness when presented with large and complex combinatorial optimization problems with uncertainties and dynamics in operation environment.

REFERENCES

Arrazola, J.M., Delgado, A., Bardhan, B.R., loyd, S.L., 2019, Quantum-inspired algorithms in practice, Quantum Physics, arXiv:1905.10415v3.

Artikis, A., Baber, C., Bizarro, P., Canudas-de-Wit, C., Etzion, O., Fournier, F, Goulart, P., Howes, A., Lygeros, J., Paliouras, G., Schuster, A., Sharfman, I., 2014, Scalable Proactive Event-Driven Decision Making, IEEE Technology and Society Magazine, 33(3): 35-41, DOI: 10.1109/MTS.2014.2345131.

Bengio, Y., Lodi, A., Prouvost, A., 2021, Machine Learning for Combinatorial Optimization: A Methodological Tour d'Horizon, European Journal of Operational Research, 290(2): 405-421.

Blum, C., Puchinger, J., Raidl, G., Roli, A., 2011, Hybrid metaheuristics in combinatorial optimization: A survey, Appl. Soft Comput., 11(6): 4135-4151, DOI: 10.1016/j.asoc.2011.02.032.

Borissova, D., Keremedchiev, D., 2019, Product configuration design via group decision making and combinatorial optimization, Comptes rendus de l'Acad´emie bulgare des Sciences, 72(9): 1251-1261, DOI: 10.7546/CRABS.2019.09.13.

Cruz, C., González, J., Pelta, D., 2011, Optimization in dynamic environments: A survey on problems, methods and measures, Soft Comput., 15: 1427-1448, DOI: 10.1007/s00500-010-0681-0.

Duan, L.-M., Raussendorf, R., 2005, Efficient Quantum Computation with Probabilistic Quantum Gates, Physical Review Letters, 95(8): 080503, DOI: 10.1103/PhysRevLett.95.080503

Ermon, S., Xue, Y., Conrad, J., Gomes, C., Selman, B., 2012, Combinatorial Decision Making in Complex, Uncertain, and Highly Dynamic Environments, 3rd International Conference on Computational Sustainability.

Fellous-Asiani, M., Chai, J.H., Whitney, R.S., Auffèves, A., Ng, H.K., 2020, Limitations in quantum computing from resource constraints, Quantum Physics, arXiv:2007.01966v1.

Fernandes Muritiba, A.E., 2010, Algorithms and Models for Combinatorial Optimization Problems, PhD Dissertation in Automation and Operations Research, Studiorum University of Bologna., DOI: 10.6092/unibo/amsdottorato/2897.

Grover, L.K., 1996, A fast quantum mechanical algorithm for database search. Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pp. 212-219.

Gao, X., Du, H., Han, M., 2017, Combinatorial Optimization and Applications, Proceedings of COCOA 2017, LNCS 10627, Springer.

Han, K.H., Kim, J.H., 2000, Genetic quantum algorithm and its application to combinatorial optimization problem, IEEE Proceedings of the 2000 Congress on Evolutionary Computation, Vol. 2, pp. 1354-1360.

Han, K.H., Kim, J.H., 2002, Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. IEEE transactions on evolutionary computation, vol. 6, no. 6, pp. 580-593.

Jin, Y., Wang, H., Chugh, T., Guo, D., Miettinen, K., 2019, Data-Driven Evolutionary Optimization: An Overview and Case Studies, IEEE Transactions on Evolutionary Computation, 23(3): 442-458.

Jouglet, A., Carlier, J., 2011, Dominance rules in combinatorial optimization problems, European Journal of Operational Research, 212(3): 433-444.

Mazyavkina, N., Sviridov, S., Ivanov, S., Burnaev, E., 2020, Reinforcement Learning for Combinatorial Optimization: A Survey, arXiv:2003.03600v3.

Moore, M.P., Narayanan, A., 1995, Quantum-inspired computing, Technical Report, Department of Computer Science, University Exeter, Exeter, UK.

Mukelabai, M., Nešić, D., Maro, S., Berger, T., Steghöfer, J.-P., 2018, Tackling combinatorial explosion: a study of industrial needs and practices for analyzing highly configurable systems, Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, pp. 155-166, DOI: 10.1145/3238147.3238201.

Neumann, F., Witt, C., 2010, Combinatorial Optimization and Computational Complexity, In: Bioinspired Computation in Combinatorial Optimization, Natural Computing Series. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-16544-3_2.

Patvardhan, C., Bansal, S., Srivastav, A., 2015, Quantum-inspired evolutionary algorithm for difficult knapsack problems. Memetic Computing, vol. 7, no. 2, pp. 135-155.

Sengupta, S., Chakraborti, T., Sreedharan, S., Vadlamudi, S.G., Kambhampati, S., 2017, RADAR-A Proactive Decision Support System for Human-in-the-Loop Planning, In AAAI Fall Symposia, pp. 269-276).

Shor, P.W., 1994, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, IEE Computer society press, Proceedings of the 35th Annual Symposium on Foundations of Computer Science.

Vogiatzis, C., Pardalos, P.M., 2013, Combinatorial Optimization in Transportation and Logistics Networks, In: Pardalos P., Du DZ., Graham R. (eds) Handbook of Combinatorial Optimization, Springer, New York, NY. https://doi.org/10.1007/978-1-4419-7997-1_63.

Weber, J.R., Koehl, W. F., Varley, J.B., Janotti, A., Buckley, B.B., Van de Walle, C.G., Awschalom, D. D., 2010, Quantum computing with defects. Proceedings of the National Academy of Sciences, 107(19), 8513-8518.

Xiong, H., Wu, Z., Fan, H., Li, G., Jiang, G., 2018, Quantum rotation gate in quantum-inspired evolutionary algorithm: A review, analysis and comparison study, Swarm and Evolutionary Computation, 42(Oct): 43-57.

Yang, H., Qiu, R., Chen, W., 2019, Smart Service Systems, Operations Management, and Analytics, Proceedings of the INFORMS International Conference on Service Science, Springer.

Zhang, G., 2011, Quantum-inspired evolutionary algorithms: a survey and empirical study. Journal of Heuristics, 17(3), 303-351.

Zhao, H., Icoz, T., Jaluria, Y., Knight, D., 2007, Application of data-driven design optimization methodology to a multi-objective design optimization problem, Journal of Engineering Design, 18(4): 343-359

DECISION SCIENCES INSTITUTE

Efficient Data Cleansing with Self-Service Data Prep Tools

Mohan Rao, PhD
Texas A&M University-Corpus Christi
mohan.rao@tamucc.edu

Navya Reddy Velverthi
University of North Texas
navyareddyvelverthi@my.unt.edu

ABSTRACT

Data analytics is rapidly growing in demand. One of the six steps in the CRISP-DM data mining process is data preparation. It is an extremely important step requiring special attention. Research findings on data mining projects report that as much as 80% of the time is spent on data preparation and only 20% on actual data analysis. The Self-service data prep tools can make the process more efficient. In this paper, we present a Gapminder application from data extraction to cleansing and developing a visual application. We also present the benefits it had for both students and the researchers.

KEYWORDS: Data analytics, big data, data science, data mining, data preparation, data prep tools, Tableau Prep Builder

INTRODUCTION

The growth of analytics is driven by the increasing availability of data, user-friendly software, and the need to make data-driven decisions for benefit of the organizations. However, the data may not be in a ready-to-use format. Essentially, data preparation is the process of gathering, structuring, organizing, and combining data. According to Kwak & Kim (2017), in the age of big data, data preparation entails more than just organizing data. It is a requirement that drives decision-making. It's necessary to make sense of multinational multibillion-dollar investment in big data analytics. It also represents a much-needed new approach to self-service data preparation, allowing business users to optimize their data for its intended use.

Through the experience working with Fortune 500 companies, Soibelman and Kim (2002) found that companies that understand the underlying business problems with data are more likely to succeed with data preparation. These organizations are aware of the issue and its implications for the business. Organizations that did not assess, comprehend, and acknowledge, on the other hand, have failed to perform well in the current competitive world (Kwak & Kim, 2017). Although everyone talks about data preparation, hardly anyone does anything about it. Following the identification of goals or the problems to be solved, data preparation is the key to resolving the problem. It can mean the difference between winning and losing, between useful insights and incomprehensible text, between informed decisions and meaningless assumptions (Soibelman & Kim, 2002).

One customer, for example, needed to use their own data to start up a customer personalization strategy. The company viewed itself as data-driven since it had created data lakes to store its customers' household survey data and now wanted to use that data to provide personalization services. While they were aware of raw data issues, they were unprepared to deal with a massive number of replicas and garbage that rendered nearly 40% of their data useless. Before they could begin working on their personalized goals, they needed to prepare and clean their data (Pyle, 1999). So, while businesses may have massive data lakes, they inevitably devolve into data dumps because the immediate thought was that 'more data is better.' This strategy is no longer effective. Data preparation and its sub-activities are required to ensure that you have usable data to work with. Otherwise, according to Pyle (1999), a lack of clean data could lead to:

- Inefficiency in operations: Teams and processes will suffer as a result of not having the right data set to work with and contribute to the goal.
- Low customer satisfaction: When a company mismanages data, it can lead to embarrassing mistakes and missed opportunities, resulting in low customer satisfaction.
- Needless costs: The consequences of mismanaged data will result in all sorts of expenses.
- Slowed growth: It's a fast-paced market out there. Users will not be able to grow if you do not rely on data. Forward-thinking businesses prioritize data optimization.
- Inaccurate insights: Nearly every single company uses data modeling for knowledge and insight and analytics. Inaccurate data that has not been subjected to a data preparation process will lead to flawed insights, the repercussions of which we are all too familiar.
- In its most basic form, data preparation assists us in understanding information in data that we can understand simply by looking at it. That's all there is to it. And that is the primary goal of this activity.

Motivation for this paper came from the idea to develop Tableau applications for performance evaluation. For a long time, these applications have been developed in Excel. A second motivation is a desire to find and use publicly available datasets from sources such as World Bank databases for student projects in data analytics. In both of these cases, the data is not in a format that could be readily used in Tableau or similar software for developing applications. It has to be cleansed and reformatted before its use. Doing this manually can be time-consuming. Self-service tools such as Tableau Prep Builder can save a lot of time.

This paper describes the process from original data, its transformation, and its use in developing applications. There are significant benefits for both learners and researchers, and consequently for organizations. The next sections include literature review, data extraction, data prepping, application and results, and conclusions.

LITERATURE REVIEW

Pérez et al. (2015) assert that it can be difficult to find a dependable self-service data preparation tool in today's market, the right solution can resolve all data management issues. To begin, the tool will be able to integrate with all business applications, allowing data to be imported. This includes both social media and CRM platforms (Salesforce and HubSpot). Modern data preparation tools can visually identify problems with your data and provide an overview of data health. The viewer must be able to see the number of typos, misspellings, and other quality issues at a glimpse, which is accurately what self-service data prep tools provide. For data cleaning, a set of strict criteria is required. These rules aid in the cleaning up of sloppy data as well as standardizing it per established standards. Furthermore, data matching is the foundation of data preparation because it is the only technique for extracting duplicates. As a result, the tool must include fuzzy matching algorithms to make sure high accuracy when it comes to data combining as well.

The next generation of Business Analytics and Business Intelligence is Self-Service Data Preparation. Self-Service Data Preparation (SSDP) makes Advanced Data Discovery available to all team members and business users, regardless of skill level or technical knowledge. Traditionally, data preparation was the primary responsibility of IT professionals and data scientists. To prepare data for analysis, one must first find, gather, and prepare the data, which includes cleaning, combining, reducing, and shaping the data (Rajagopalan & Isken, 2001). This complex process meant that the average manager or business user could not receive analysis or reports in a timely way, and the results were frequently difficult to interpret. This prompted the need for SSDP, which we will discuss in the following paragraphs.

SSDP is a collection of sophisticated tools designed for business users' ease of use and access in a self-service environment. Self-Service Data Prep empowers users to explore, manipulate, and merge new data sources without requiring the assistance of an IT staff. Its significance in the business world cannot be overstated. With self-service data prep, data analytics moves away from analysts and IT and into the hands of business users (Rajagopalan & Isken, 2001). The average business user can perform data preparation, test theories, and hypotheses by prototyping on their own and share clear, objective data with others with true self-service business intelligence and analytics solutions.

Self-Service Data Preparation enables business users to use tools without the constraints imposed by managed dashboards or standardized reporting tools. Without the assistance of programmers or data scientists, users can access complex tools in an easy-to-use environment. SSDP empowers business customers by enabling them to complete tasks, make decisions, and make recommendations with speed, agility, and accuracy (Pyle, 1999). SSDP also enables average enterprise customers to collect and organize information for use in analytics to test hypotheses, observe and share data, prepare reports, and support daily operations tasks with complete drill-down and drill-through capacity, custom alerts, and mobile access that satisfy the requirements of every team member. Users can control the data elements, volume, and timing of the analysis and reporting. You can easily publish and run flows in your server environment using Tableau Prep Conductor (Pyle, 1999). Using Tableau Server or Tableau Online, you can securely share your data sources. Make it possible for everyone in your organization to work with prepared and up-to-date data.

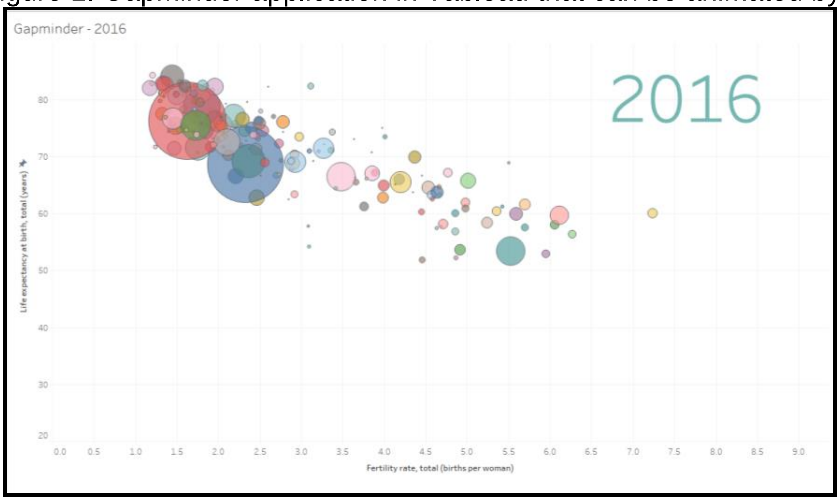
Some organizations, such as nuclear power plants and earthquake bureaus, require external data collection because their databases are so small. For example, because nuclear power plant accidents cause many environmental disasters, economic and ecological damage, and endanger people's lives, automatic surveillance and early nuclear accident detection have received a lot of attention. To reduce nuclear accidents, we need reliable knowledge for nuclear accident control. A nuclear accident database, on the other hand, frequently contains insufficient data to form reliable patterns. As a result, mining the nuclear power plant's accident database must rely on external data. Yan et al. (2003) propose a new method for distinguishing between external and internal knowledge from various data sources, and they use relevant belief knowledge to rank potential facts. They use a pre-or post-analysis to assess the relevance of recognized external data sources to the data-mining problem, rather than common data cleaning tasks like removing errors and filling missing values. This paper proposes a novel approach to data pre-processing work in the mining of data.

Businesses of all sizes can benefit from big data, even on a small scale — but many don't know where to start. For those businesses, the solution is simple: "get organized." Analyzing business data is similar to any other application of the scientific method in that it must begin with high-quality data. Data preparation necessitates cross-departmental collaboration, attention to detail, and a clear understanding of the mission or problem at hand. To do well, it will almost certainly necessitate the use of the appropriate software (Pyle,1999). However, getting it right entails extracting "free" value from data that you most likely already have, such as discovering new answers to old questions and discovering new ways to reach customers, optimize performance, reduce waste, or achieve any number of other business objectives.

DATA EXTRACTION

Finding data for student projects in Data Analytics courses is always a challenge. We came across World Bank databases from watching Hans Rosling's inspirational TED Talk presentation on Gapminder. We realized it is an excellent source. However, after extraction, the data needs to be cleansed and reformatted to develop any visualization applications. To motivate students to get data on various topics from World Bank and other similar databases, first, they needed to learn cleansing and reformatting. Therefore, an assignment was created for students to develop the Gapminder application in Tableau that can be animated by year (Figure 1).

Figure 1. Gapminder application in Tableau that can be animated by year



To do that, first, they need to extract data from the WDI database and prepare it for use in Tableau Desktop. Then the assignment sheet provides step-by-step instructions showing how to extract data, how to clean the data using Tableau Prep Builder. Students need to submit (1) Extract from WDI, (2) Output from Tableau Prep, and (3) Tableau Prep packaged file in a zip file.

Gapminder application begins with the data extraction from the WDI database on the World Bank website. The step-by-step instructions include the following:

Go to the World Bank website: https://databank.worldbank.org/source/world-development-indicators#. World Development Indicators is the first one of the 78 databases. For the Gapminder application, we need only three variables (Series names or Indicator names):

1. Fertility rate, total (births per woman)
2. Life expectancy at birth, total (years)
3. Population, total

Once we go through the selection of the data from the WDI database, the downloaded file may look like the following Figure 2. It will have about 64 columns depending on the years of data, and 652 rows, including the header.

Figure 2. Sample file from World Bank WDI database

	A	B	C	D	E	F	G	H	I	J
1	Country Name	Count	Series Name	Series Code	1960 [YR19	1961 [YR19	1962 [YR19	1963 [YR19	1964 [YR19	1965 [Y
2	Afghanistan	AFG	Fertility rate, total (births per	SP.DYN.TFRT.IN	7.45	7.45	7.45	7.45	7.45	7
3	Afghanistan	AFG	Life expectancy at birth, total	SP.DYN.LE00.IN	32.446	32.962	33.471	33.971	34.463	34.
4	Afghanistan	AFG	Population, total	SP.POP.TOTL	8996973	9169410	9351441	9543205	9744781	9956
5	Albania	ALB	Fertility rate, total (births per	SP.DYN.TFRT.IN	6.489	6.401	6.282	6.133	5.96	5

Data | Series - Metadata

The above data becomes the input to a data prep tool like Tableau Prep Builder so that it is cleaned and reformatted for application development using a tool like Tableau Desktop.

DATA PREPPING

Once the data is extracted, the next step is to clean it and reformat it so that it can be used to develop analytic applications in software such as Tableau Desktop. For cleaning and reformatting, there is another set of instructions (not shown here) provided to the students so that they will get the output shown in Figure 3 from Tableau Prep Builder.

Figure 3. Sample output from Tableau Prep

	A	B	C	D	E	F	G
1	Number of Records	Population, total	Life expectancy at birth, total (years)	Fertility rate, total (births per woman)	Year	Country Name	Country Code
2	1	8996973	32.446	7.45	1960	Afghanistan	AFG
3	1	1608800	62.283	6.489	1960	Albania	ALB
4	1	11057863	46.141	7.524	1960	Algeria	DZA
5	1	20123			1960	American Samoa	ASM
6	1	13411			1960	Andorra	AND
7	1	5454933	37.524	6.708	1960	Angola	AGO
8	1	54131	61.968	4.425	1960	Antigua and Barbu	ATG

GapminderOutputfrom Prep

We can see the difference between Figures 2 and 3. The data in this format in Figure 3 can be input into Tableau Desktop for developing the Gapminder application shown in Figure 1.

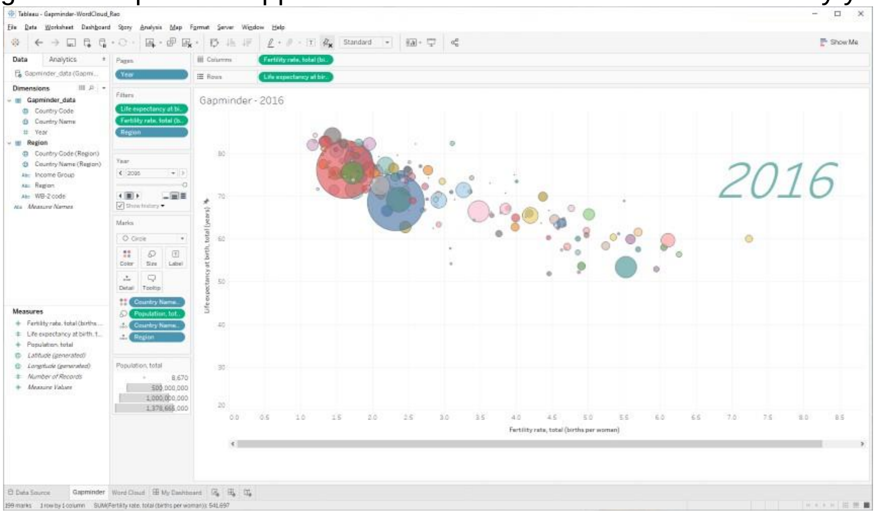
APPLICATION AND RESULTS

The instructions for developing the Gapminder application are as follows:

- Connect to Gapminder_data.xlsx (data shown in Figure 3)
- Connect Gapminder_data and Region with Left join
- Fertility Rate (x) vs. Life expectancy (y) Scatterplot. Uncheck Aggregation. Pivot, if necessary. Adjust y-axis to min 20 fixed.
- Year to Pages
- Population to Size. Increase size.
- Country Name (Region) to Color. Adjust colors, if needed.
- Filter Region. Exclude Null. Exclude any other null values.
- Choose Circle Mark. Lower Opacity and choose an outline.
- Annotate Area. Insert Page Name. Size to 72. Adjust how the year shows. Format>Remove Shading.

These steps lead to the following figure that can be animated by year.

Figure 4. Gapminder application in Tableau that can be animated by year



Based on this experience, students were able to use several open-source datasets, cleanse, reformat and develop applications for data analytic projects.

CONCLUSIONS

In this paper, we showed the importance of data preparation for any data analytic application. We have also presented a Gapminder application starting with extracting data from the WDI database (Figure 2), cleaning and reformatting it (Figure 3) using Tableau Prep Builder, and then developing an application using Tableau Desktop (Figure 4). This experience has helped the students and authors to use various datasets, cleanse, reformat, and develop applications.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE
A New General Searls Estimation Procedure of Population Mean of Primary Variable

Dinesh K. Sharma
University of Maryland Eastern Shore, USA
Email: profdksharma@gmail.com

S. K. Yadav[1]
Babasaheb Bhimrao Ambedkar University, India
Email: drskystats@gmail.com
[1]Corresponding Author

Padmeswar Doley
Shri Ram College of Commerce, University of Delhi, India
Email: padmeswar.doley@srcc.edu

Shakti Kumar
Jawahlal Nehru University, India
Email: shaktinov30@yahoo.co.in

ABSTRACT

This study aimed to determine the population mean of the main variable without raising the survey expense. We suggest a new estimator for the population mean of the main variable in this paper that takes advantage of readily available data on the population median of the primary variable. We investigate the bias and mean squared errors (MSE) of the introduced family of estimators up to an order one approximation. For the optimal values of the characterizing scalars, the introduced estimator has the lowest MSE. The introduced class is compared with competing estimators that employ known auxiliary parameters obtained at an additional survey cost. The efficiency conditions are theoretically determined and empirically validated. Finally, the results are displayed in a tabular format. For both actual and simulated populations, simulation and computation were carried out using R programming language codes.

KEYWORDS: Study variable, Median, Bias, MSE, Percentage Relative Efficiency (PRE).

INTRODUCTION

A common theme in the literature is that the use of supplementary data enhances the sampling design. The auxiliary data is derived from the auxiliary variable (X), having a high correlation with the primary variable (Y). The ratio and product methods, which are highly positively and negatively correlated, are used to estimate the population mean ($\bar{Y}$). Cochran (1940) utilized the positively correlated X and suggested the standard ratio estimator of $\bar{Y}$ and outperformed the sample mean estimator, which is thought to be the most appropriate estimator. Furthermore, Sisodia and Dwivedi (1981) utilized a known population coefficient of variation X and introduced an elevated estimator $\bar{Y}$. Singh and Kakran (1993), and Kadilar and Cingi (2004) proposed the known population coefficient of kurtosis of X for enhanced estimation of $\bar{Y}$. Singh and Tailor (2003) suggested an improved estimator of $\bar{Y}$ by utilizing the known population coefficient of correlation between X and Y . Yan and Tian (2010) introduced an enhanced estimator of $\bar{Y}$ by using the known population coefficient of skewness of X. Despite the above significant contributions, numerous authors have introduced various elevated estimators of $\bar{Y}$ based on various known auxiliary parameters.

In this paper, we consider a new estimator for improved estimation of $\bar{Y}$ using known median (M) of Y, which is motivated by several authors in literature. As a result, we may be able to estimate the population mean more closely to the true population mean, and policies may be formed more effectively. We investigate the proposed estimator's various properties and compare it theoretically and empirically with competing estimators to determine its efficiency over the estimators in competition. The entire study has been divided into various sections, including a review of existing estimators using auxiliary information, the proposed estimator, theoretical efficiency comparison, empirical study, results, and discussion, and the manuscript concludes with a conclusion.

REVIEW OF EXISTING ESTIMATORS

In this section, different estimators of $\bar{Y}$ has been presented along with their variance/MSEs up to the first order of approximation. Table 1 below shows different estimators of $\bar{Y}$ under consideration with their MSEs.

Table 1: Estimators by different authors and their MSE

S. No.	Author(s)	Variance/MSE
1.	Anonymous for sample mean	$V(t_0) = \gamma \bar{Y}^2 C_y^2$
2.	Watson (1937)	$V(t_{reg}) = \gamma \bar{Y}^2 C_y^2 (1-\rho^2)$
3.	Cochran (1940)	$MSE(t_r) = \gamma \bar{Y}^2 (C_y^2 + C_x^2 - 2C_{yx})$
4.	Goodman and Hartley (1958)	$V(t_1) = \gamma \bar{Y}^2 (C_y^2 + C_x^2 - 2C_{yx})$
5.	Chakrabarty (1979)	$MSE_{\min}(t_2) = \gamma \bar{Y}^2 C_y^2 (1-\rho^2)$
6.	Sahai and Ray (1980)	$MSE_{\min}(t_3) = \gamma \bar{Y}^2 C_y^2 (1-\rho^2)$
7.	Sisodia and Dwivedi (1981)	$B(t_4) = \gamma \bar{Y} (R_4^2 C_x^2 - R_4 C_{yx})$
8.	Bahl and Tuteja (1991)	$MSE(t_5) = \gamma \bar{Y}^2 (C_y^2 + \frac{C_x^2}{4} - C_{yx})$
9.	Upadhyaya and Singh (1999)	$B(t_i) = \gamma \bar{Y} (R_i^2 C_x^2 - R_i C_{yx})$, $i = 6, 7$
10.	Kadilar and Cingi (2003)	$MSE(t_8) = \gamma \bar{Y}^2 (C_y^2 + 4C_x^2 - 4C_{yx})$
11.	Singh (2003)	$MSE(t_9) = \gamma \bar{Y}^2 (C_y^2 + R_9^2 C_x^2 - 2R_9 C_{yx})$
12.	Singh and Tailor (2003)	$MSE(t_{10}) = \gamma \bar{Y}^2 (C_y^2 + R_{10}^2 C_x^2 - 2R_{10} C_{yx})$
13.	Singh *et al.* (2004)	$MSE(t_{11}) = \gamma \bar{Y}^2 (C_y^2 + R_{11}^2 C_x^2 - 2R_{11} C_{yx})$
14.	Yan and Tian (2010)	$MSE(t_i) = \gamma \bar{Y}^2 (C_y^2 + R_i^2 C_x^2 - 2R_i C_{yx})$, $i = 12, 13, 14, 15$
15.	Al-Omari (2012)	$MSE(t_{16}) = \gamma \bar{Y}^2 (C_y^2 + R_{16}^2 C_x^2 - 2R_{16} C_{yx})$,

16.	Subramani & Kumarpandiyan (2012a, b)	$MSE(t_i)=\gamma\bar{Y}^2(C_y^2+R_i^2C_x^2-2R_iC_{yx})$, $i=17,18,19$
17.	Swain (2014)	$MSE(t_{20})=\gamma\bar{Y}^2(C_y^2+\frac{C_x^2}{4}-C_{yx})$
18.	Yadav and Mishra (2015)	$MSE(t_{21})=\gamma\bar{Y}^2(C_y^2+\alpha^2\frac{C_x^2}{4}-\alpha C_{yx})$
19.	Jerajuddin and Kishun (2016)	$MSE(t_{22})=\gamma\bar{Y}^2(C_y^2+R_{22}^2C_x^2-2R_{22}C_{yx})$
20.	Subramani (2016)	$MSE(t_{23})=\gamma\bar{Y}^2(C_y^2+R_{23}^2C_m^2-2R_{23}C_{ym})$
21.	Yadav *et al.* (2017)	$MSE(t_{24})=\gamma\bar{Y}^2(C_y^2-\frac{C_{ym}^2}{C_m^2})$
22.	Soponviwatkul and Lawson (2017)	$MSE_{\min}(t_i)=\gamma\bar{Y}^2C_y^2(1-\rho^2)$, $i=25,26$
23.	Ijaz and Ali (2018)	$MSE_{\min}(t_i)=\gamma\bar{Y}^2C_y^2(1-\rho^2)$, $i=27,28$
24.	Yadav *et al.* (2019)	$MSE_{\min}(t_{29})=\gamma\bar{Y}^2C_y^2(1-\rho^2)$

Where,
$\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}y_i$ is the population mean of y, $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ is the sample mean of y, $C_y=\frac{\sigma_y}{\bar{Y}}$ is the coefficient of variation of y, $\sigma_y=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{Y})^2}$ is the standard deviation y, $\gamma=\left(\frac{1}{n}-\frac{1}{N}\right)$ with n as sample size and N as population size, $\bar{X}=\frac{1}{N}\sum_{i=1}^{N}x_i$ is the population mean of x, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ is the sample mean of x, $C_x=\frac{\sigma_x}{\bar{X}}$ is the coefficient of variation of x, $\sigma_x=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{X})^2}$ is the standard deviation of x, $C_{yx}=\rho C_yC_x$, ρ is the coefficient of correlation between y and x, $R_4=\frac{\bar{X}}{\bar{X}+C_x}$, $R_6=\frac{\bar{X}C_x}{\bar{X}C_x+\beta_2}$, $R_7=\frac{\bar{X}\beta_2}{\bar{X}\beta_2+C_x}$, $R_9=\frac{\bar{X}\beta_2}{\bar{X}\beta_2+S_x}$, $R_{10}=\frac{\bar{X}}{\bar{X}+\rho}$, $R_{11}=\frac{\bar{X}}{\bar{X}+\beta_2}$, $R_{12}=\frac{\bar{X}}{\bar{X}+\beta_1}$, $R_{13}=\frac{\bar{X}\beta_1}{\bar{X}\beta_1+\beta_2}$, $R_{14}=\frac{\bar{X}C_x}{\bar{X}C_x+\beta_1}$, $R_{15}=\frac{\bar{X}\beta_2}{\bar{X}\beta_2+\beta_1}$, $R_{16}=\frac{\bar{X}}{\bar{X}+Q_3}$, $R_{17}=\frac{\bar{X}}{\bar{X}+Q_r}$, $R_{18}=\frac{\bar{X}}{\bar{X}+M_d}$, $R_{19}=\frac{\bar{X}C_x}{\bar{X}C_x+M_d}$, $R_{22}=\frac{\bar{X}}{\bar{X}+n}$, $R_{23}=\frac{\bar{Y}}{M}$, $R_{29}=\frac{a.b\bar{X}}{a.b\bar{X}+c.d}$.

PROPOSED ESTIMATOR

Under this section, we have suggested a generalized ratio type estimator of $\bar{Y}$ utilizing the known population median of the main variable and have studied the bias and MSE up to the first order of approximation. The estimator, bias, and MSE of the suggested estimator, respectively are given as,

$$t_p = \kappa_1 \bar{y} + \kappa_2 \log \frac{m}{M}$$

$$B(t_p) = \bar{Y}\left[\kappa_1 - 1 + \kappa_2 \left\{ \frac{\bar{M} - M}{M} + C_{ym} - \frac{C_m^2}{2}) \right\}\right]$$

$$MSE_{\min}(t_p) = \bar{Y}^2 [1 + P - 2Q]$$

Where,

$$P = \left\{ \frac{A(BC - A_1)^2 + A_1(B - AC)^2 + 2B(BC - A_1)(B - AC)}{(B^2 - A_1 A)} \right\}$$ and

$$Q = \left\{ \frac{(BC - A_1) + (B - AC)C}{(B^2 - A_1 A)} \right\}$$

EMPIRICAL STUDY

In this section, we have compared different estimators under comparison and have calculated the variance and MSE of these estimators. The parameters of the population under consideration are presented in Table 2.

Table 2: Parameters of the Natural Population
$N = 52$, $n = 3$, $\gamma = 0.3141$, ${}^N C_n = 22100$, $\bar{Y} = 14.7211$, $\bar{X} = 0.4623$, $M = 10$, $\bar{M} = 12.0119$, $\rho = 0.8046$, $S_y^2 = 190.8668$, $S_x^2 = 0.1568$, $S_{yx}^2 = 4.4012$, $C_y^2 = 0.8807$, $C_x^2 = 0.7335$, $C_{yx} = 0.3083$, $\beta_1 = 8.1029$, $\beta_2 = 14.1463$, $\beta = 28.0769$, $Q_1 = 0.4040$, $Q_3 = 0.5050$, $Q_r = 0.2525$, $C_m^2 = 0.3774$, $C_{ym}^2 = 0.2536$

The biases, MSEs of the suggested and the competing estimators, and the percentage relative efficiency of the suggested estimator over the competing estimators under a simple random sampling scheme are presented in Table 3.

Table 3: MSE of various estimators under comparison

Estimator	MSE	Estimator	MSE	Estimator	MSE
t_0	59.9518	t_9	83.5876	t_{20}	27.1859
t_{reg}	21.1379	t_{10}	21.1637	t_{21}	28.6051
t_r	21.8372	t_{11}	34.4728	t_{22}	28.4113
t_1	21.8372	t_{12}	57.2156	t_{23}	14.6939
t_2	21.8372	t_{13}	55.3452	t_{24}	14.2061
t_3	21.1379	t_{14}	43.7074	t_{25}	15.9282
t_4	21.1379	t_{15}	57.3737	t_{26}	49.0862
t_5	35.2235	t_{16}	30.5886	t_{27}	21.1379
t_6	28.4113	t_{17}	21.4937	t_{28}	21.1379
t_7	57.5916	t_{18}	29.2776	t_{29}	21.1379
t_8	21.1382	t_{19}	23.8936	t_p	**11.2095**

CONCLUSION

In the present paper, we have suggested a new general ratio type estimator to estimate $\bar{Y}$ using the known population median of Y. The bias and MSE of the proposed estimator are studied up to the first order of approximation. The suggested estimator is compared with the competing estimators of $\bar{Y}$ using and without using auxiliary information. From Table 3, it may be observed that the suggested estimator has the least MSE among the class of estimators in competition. Thus, the introduced estimator may be utilized for the enhanced estimation of $\bar{Y}$ in different areas of applications.

REFERENCES

Al-Omari, A.I. (2012). Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling, *Statistics and Probability Letters*, 82, 1883-18890.

Bahl, S., & Tuteja, R.K. (1991). Ratio and product type exponential estimator, *Information and Optimization Sciences*, XII, I, 159-163.

Chakrabarty, R.P. (1979). Some ratio estimators, *Journal of the Indian Society of Agricultural Statistics*, 31, 1, 4957.

Cochran, W.G. (1940). The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce, *The Journal of Agric. Science*, 30, 262-275.

Haq, A., Khan, M., & Hussain, Z. (2017). A new estimator of finite population mean based on the dual use of the auxiliary information, *Communications in Statistics-Theory and Methods*, 46, 9, 4425-4436.

(A complete list of references is available upon request)

DECISION SCIENCES INSTITUTE

An approach to preventing conflict: unpacking the standing neutral

Dan Bumblauskas, Ph.D.
University of Northern Iowa
Email: daniel.bumblauskas@uni.edu

Kate Vitasek
University of Tennessee Knoxville
Email: kvitasek@utk.edu

Jim Groton
Eversheds Sutherland (Retired)
Email: jimgroton@eversheds-sutherland.com

ABSTRACT

Contractual relationships are vulnerable to the damage caused by friction in relationships, particularly when this turns into a formal dispute. If negotiations fail, the next step is calling in a mediator, and perhaps ending up in arbitration or the court system. This paper argues for the proactive use of a "Standing Neutral" – a trusted, independent expert advisor (or a panel of three advisors) chosen by the contracting parties to help during their contractual relationship. Rather than being reactive to disputes, an organization can be proactive by a standing neutral as chosen by the contracting parties to help during their engagement.

Keywords: Standing Neutral, Third Parties (e.g., ADR, arbitration, mediation), Group Bargaining and Negotiation (i.e., multiparty), Negotiation Process & Outcomes (e.g., psychological, economic, relational)

INTRODUCTION

Contracts are inherently incomplete. No lawyer has yet crafted the perfect contract that will anticipate every eventuality. Problems and unexpected events are always around the corner. Long-term contractual relationships are especially vulnerable to the damage caused by friction in relationships, particularly when this friction turns into a formal dispute. For far too many organizations, the need to engage in conventional conflict resolution does not begin until the parties have experienced real pain. By that time, they have blamed each other for their troubles. Unfortunately, this usually means that internal escalations have failed and the parties' relationship has reached a breaking point which can lead to calling on their respective lawyers who are not typically incentivized or instinctively inclined to resolve conflicts constructively in the way best suited to the preservation of the relationship. And if typical negotiations fail, the next step is calling in a mediator, and perhaps eventually ending up in arbitration or the court system.

The fact is that most disputes start small, typically stemming from misaligned or unclear expectations. This article argues for the proactive use of a "Standing Neutral" – a trusted, independent expert advisor (or a panel of three advisors) chosen by the contracting parties to help parties during their contract relationship. A classic Standing Neutral process can best be described as a quick, informal, flexible, adaptable, non-adversarial, neutral, expert, and preferably nonbinding process for preventing and achieving the earliest possible solution to problems and potential disputes. The modern Standing Neutral is even more proactive, putting even more focus on avoiding problems to prevent disputes.

This article will help you understand the why, what and how of using a Standing Neutral for preventing and managing conflicts. The bottom line? It is your bottom line. Using a Standing Neutral is more effective and efficient for governing today's modern commercial relationships.

RESEARCH MOTIVATION FOR A STANDING NEUTRAL

Contracts can be similar to buying a new pair of shoes; it is often great at first - but sometimes friction occurs. Small misalignments are like the rub from that brand-new pair of shoes which can be unpleasant and turn into a blister. Left unchecked what starts as friction or misaligned interests can turn into a full-blown dispute – or worse – end up in court. While the vast majority of conflicts avoid litigation, the time and cost associated with traditional negotiation, mediation and arbitration can be protracted and expensive. Even if the issue does not go to a formal dispute, the friction causes lost opportunity, value leakage and transaction costs: what Oliver Williamson calls Transaction Cost Economics.

The simple fact is that friction should be expected in any complex contract. Why? In the words of Nobel Laureate Oliver Williamson: "all complex contracts will be incomplete. There will be errors, omissions, and the like (Williamson, 2008)." The very nature of complex contracts means it is impossible to predict every 'what if' scenario given today's global and dynamic business environment.

Another Nobel Laureate – Oliver Hart – echoes Williamson's sentiments regarding incomplete contracts. Hart's research with John Moore suggests you really should not blame 'the other guy' for what may seem like opportunistic behaviors, rather it stems from what Hart calls *shading* (Hart and Moore, 2008). Shading is not opportunistic behavior, but *retaliating behavior* in which a party stops cooperating, ceases to be proactive, or makes countermoves because of disappointment. Shading happens when a party doesn't get the outcome they expect from the deal and feels the other party is to blame for it or does not act reasonably by helping to mitigate the losses.

To provide another example, think about the highly contentious scope-creep debate. A supplier does a business case based on the information shared during a competitive bid process. They estimate they will make 15% profit margin. If demand is lower than expected or there is extra work not anticipated (e.g., scope-creep), the supplier will have lower than expected profit. This disappointment will cause the supplier to justify asking for scope changes. And if the buying organization makes it difficult to get a contract changes through, the supplier may be tempted to reduce service levels or replace the expensive A-team with the less costly C-team. In short – each party's action leads to the other party's reaction, creating the negative tit-for-tat cycle. One disappointment leads to another and the vicious cycle begins. The problem is so systemic in large and complex deals it is sometimes called the *death spiral* because once the cycle starts it often ends with an incumbent supplier losing the work to a competitor during the next bid cycle.

Sadly, the root cause is often not opportunism, but disappointment based on the expectations that the parties have.

The concept of shading makes sense, especially with complex deals. In complex deals, a contract will always be incomplete with gaps, errors or omissions opening the door for shading behavior after the contract is signed. Traditional contracts rarely contain proactive alignment mechanisms to avoid disappointments.

In far too many situations, the need to begin a process for dealing with disputes isn't recognized until after those disappointments have led to real pain and frustration. The reality is many issues do not resolve themselves easily and they drag on. However, most issues can be prevented, or at least resolved, while they are still small. This paper argues using the more preventive and proactive approach of a Standing Neutral to collaboratively resolve any differences in "real time" when any issues or misalignment is still small. Figure 1 provides an overview of the standing neutral definition.

Figure 1. What is a Standing Neutral?

A Standing Neutral is an innovative and promising improvement on traditional Alternative Dispute Resolution (ADR) techniques. A Standing Neutral process uses a highly qualified and respected expert, pre-selected - or "standing" - neutral who helps parties resolve issues throughout the life of a relationship. The classic Standing Neutral plays a facilitation role to help the parties see each other's perspectives and, when appropriate, provides a non-binding recommendation. Some companies use variations such as a Standing Mediator or even a Standing Arbitrator. Others have found including a Standing Neutral upfront in the actual design and creation of contract can lead to significant value and a more fair, balanced, "win-win" contract (Vitasek and Manrodt, 2012). A modern approach to the Standing Neutral concept is to engage the neutral early on, facilitating proactive and constructive dialogues and day-to-day discussions, with the aim to provide continuous alignment and prevent issues altogether.

Organizations adopting Standing Neutrals proactively acknowledge the reality that no relationship is perfect, and no contract can cover every eventuality. Errors, omissions and ambiguities can result in misinterpretation. Small things such as "does this idea count for gainsharing?" or interpretation of performance data "did the supplier score 3 or 4 on the scorecard?" cause frustration – especially for suppliers who may feel they do not have a voice. In addition, complex contracts operating in dynamic environments require frequent adjustments.

Unpacking the Standing Neutral

The time is ripe for commercial relationships to benefit from demonstrated successful experience with the Standing Neutral method. While there are many Alternative Dispute Resolution (ADR) techniques, the focus of this article is on the Standing Neutral because it is probably the least widely understood, yet most useful, of all the ADR techniques. It is also geared to not just the legal community, which traditionally focuses on court proceedings and ADR, but also to a broader business community that includes business professionals in joint ventures, long-term business arrangements, and outsourcing arrangements.

LITERATURE REVIEW

Alternative dispute resolution techniques, and use of third parties in relationships, are not new. But what is new is the emerging emphasis on shifting away from dispute resolution to dispute prevention processes – especially as they are related to using what is known as a Standing Neutral.

The Rise of Alternative Dispute Resolution for Solving Conflicts

Judicial (court) systems for resolving disputes have been in effect throughout civilized history. And almost as long as courts have been used, individuals and organizations have sought simpler, more efficient and more cost-effective means to deal with disputes – processes known today as Alternative Dispute Resolution (ADR) techniques. While modern ADR methods have only been in place for 40 years, one could argue the roots of ADR date all the way back to a decree issued by the Chinese emperor, Kang-Hsi (1654-1722).

Michael McManus and Briana Silverstein (2011) document the history of ADR in "Brief History of Alternative Dispute Resolution in the United States." Their research revealed formal ADR techniques date as far back as the Norman Conquest which allowed for a local and highly respected layperson to conduct informal, quasi-adjudicatory settings in their communities rather than use a more formal King's court.

The concept of using alternatives to a King's court was expanded more formally in the early trade guilds that sought to enforce standards of quality, performance and marketplace behavior. Many of those systems continue today in commercial markets such as the diamond market and the textile industry (Wolaver, 1934).

Pilgrims brought the concept of ADR to the United States "preferring to use their own mediation process to deal with community conflicts." When disagreements occurred, members of the community would hear claims, determine fault, assess damages, and ensure that the parties reconciled with one another (McManus and Silverstein, 2011). Mediation was formally institutionalized in the U.S. in 1898 when Congress, following initiatives begun a few years earlier in Massachusetts and New York, authorized mediation for collective bargaining disputes (Ibid). In 1925 Congress passed the Federal Arbitration Act, which included express authorization for courts to enforce arbitration awards (Ibid).

The modern terms "Alternative Dispute Resolution," and "ADR," were coined as a result of the first Pound conference in 1976, which promoted the use of mediation and arbitration as adjuncts to the traditional legal system. The Pound Conference marked the beginning of a formal movement which encouraged the business world to actively embrace out-of-court processes for managing conflict. The movement attempted to move the dispute resolution process farther "upstream," closer to the origins and sources of disputes. The Global Pound Conference series was inspired by Harvard law professor Roscoe Pound and the name of a 1976 conference named for him; it was an impetus for the growth in the popularity of arbitration and mediation in the USA.

The 1980's were a decade of increased interest and use of ADR. In 1983 the Center for Public Resources (now more aptly named the International Institute for Conflict Prevention and Resolution, or CPR) was established as a think tank for the improvement of ADR processes. In a 1984 address to the American Bar Association, then–Supreme Court Chief Justice Warren Burger advocated for lawyers to increase their use of ADR. He acknowledged that while trials are the only way to resolve some disputes, the legal system is adversarial, too costly, painful, destructive, and inefficient to effectively manage all disputes (Dunnewold, 2009).

The Advent of Proactive Prevention Practices

Even before ADR was taking hold, the business attorney Louis M. Brown (1950) argued that new ideas and innovative processes for the *anticipation of conflict* and *dispute prevention* were needed. He called this "preventive law." Brown's work sparked an interest in the formal study of preventive conflict approaches – most notably in the construction industry, which is notorious for costly disputes that can have dire consequences on the timeliness and success of a project. Brown famously argued, "It usually costs less to avoid getting into trouble than to pay for getting out of trouble."

One of the earliest known preventive practices dates back the early 1900s when the American Institute of Architects established a system for resolving construction project disputes between project owners and contractors that designated the architect as the initial judge of the contractor's performance. In case of a dispute over the architect's decision, the process called for a prompt appeal to an *ad hoc*, one-issue arbitration before an expert construction industry arbitrator. Typically, there was usually no shortage of qualified individuals who could serve as arbitrators on short notice.

The American Institute of Architects system emphasized timeliness – which was crucial for fast-moving construction projects where delays can be costly and have significant negative impact. Further, the easy availability of an immediate decision in arbitration encouraged architects to act fairly and with integrity, usually resulting in mutual acceptance of architect's decision without an appeal, thus avoiding and preventing any dispute.

The practice of using preventive techniques in the construction industry was expanded in 1975 when a group of innovative construction practitioners conceived of the idea of a "dispute review board" of geological engineers to immediately solve difficult rock and soil problems on a major tunneling construction project (Mathews et al., 1996, p. 10). By the mid-1980s owners and contractors on major industrial projects further expanded on the concept by developing long-term "trusting" alliances to achieve greater efficiency and cost savings, processes which they called "strategic partnering," which evolved into "project specific partnering (Ibid)." During the same period other advances also emerged such as using "financial incentives to encourage cooperation (Construction Industry Institute, 1986)," and the concept of realistically allocating risks on construction projects, to achieve maximum efficiency (Construction Industry Institute (1988).

A common characteristic of these contractual preventive processes – in contrast to conventional ADR "resolution" processes – is that they proactively address problems and potential disputes before they morph into intractable disputes, rather than reactively deal with disputes after they have occurred.

The 1990's began more formal study into preventive techniques. The Construction Industry Institute (CII) led the pack with significant research between 1991 and 1994 that validated the utility of prevention processes being piloted. The CII added to the body of knowledge by suggesting the use of a "disputes potential index" to identify potential sources of trouble on construction projects (Diekmann, 1994). CII also demonstrated that the dispute review board concept could be expanded to multi-disciplinary projects such as high-rise office buildings and not just projects involving single technical disciplines (Vorster, 1993).

A key part of the CII's research contribution was the formal recognition of a critical distinction between "preventive" techniques and "resolution" techniques in the dispute resolution. This distinction was known as the "continental divide of dispute resolution" and is the point where parties lose control, and the process moves from proactive *prevention* to reactive *resolution* which "is neither timely nor cheap and is seldom satisfactory (Vorster, 1993, p. 10)." The terms are defined as:

- "Preventive" techniques: processes that enable the parties (and persons in privity with the parties) to keep control of their disagreement and avoid conventional dispute resolution
- "Resolution" techniques: processes through which "outsiders" or "strangers" to the disagreement seek to resolve a dispute
  During the same timeframe the American Arbitration Association (Groton, 1990) and the

Center for Public Resources (1991) also advanced the understanding of preventive approaches when they classified these approaches into a spectrum or continuum of progressive dispute prevention processes which could be combined into graduated processes or "systems" to provide contracting parties a full range of dispute prevention and resolution alternatives (Costantino and Sickles Merchant, 1996).

This early work in effect moved the dispute resolution process even further "upstream," embracing proactive processes that anticipate, deal with and prevent problems and potential disputes at the source, before they must be subjected to traditional, expensive, time-consuming and potentially relationship-damaging dispute resolution.

One of the most recent preventative techniques is the use of data analytics to search for early warning" indicators. This idea emerged in 2012 when data analytics experts began to analyze electronic data files (e.g., documents, emails, texts) to detect patterns that might indicate a potential risk. If a risk is detected, the suspected data can be analyzed and addressed before a potential problem develops into a serious concern. For example, if a potentially risky problem appears in yesterday's emails, inside counsel may decide to conduct an internal investigation today to confirm or deny the "early warning." If confirmed, and since the text at issue has been surfaced in near real-time, risks can be addressed internally before they develop into real problems, disputes, or litigation (Brestoff, 2012; Brestoff and Inmon, 2015).

While the construction industry was having early successes with preventing practices, other industries were still stuck in “resolution” – which was growing more complicated and costlier. Carver and Vondra (1994) critiqued ADR in the aptly titled article “Alternative Dispute Resolution: Why It Doesn’t Work and Why It Does” which stated: ”The bad news is that ADR as currently practiced too often mutates into a private judicial system that looks and costs like the litigation it’s supposed to prevent.” Cornell Law School reports this has only worsened over the years (Legal Information Institute, 2017).

The good news is there is increased interest from the academic and business communities in expanding “preventive” ADR techniques outside of the construction industry. While arbitration and mediation are still the most well-known and used ADR techniques today, the concept of ADR of has grown to mean any method of resolving disputes without resorting to litigation in a courtroom (Legal Information Institute, 2017). In addition, using preventive processes has expanded beyond the construction industry to many other kinds of business relationships (International Institute for Conflict Prevention & Resolution, 2010).

Preventive practices got a boost when the International Mediation Institute (IMI) organized a follow up to the original Pound Conference in 2016/2017 to evaluate the state of dispute resolution 40 years after the first Pound Conference of 1976. The aforementioned series of conferences (known as the Global Pound Conferences) were held in 29 cities around the world and brought together thousands of users, providers and advisors to discuss the future direction of ADR. During the events delegates were asked to vote on which dispute resolution processes should be prioritized to improve dispute resolution. In the overall cumulative voting, the delegates - by a substantial margin - voted for “pre-dispute or pre-escalation processes to prevent disputes,” over all other dispute resolution processes. The title of the conference series was "Shaping the Future of Dispute Resolution & Improving Access to Justice.” After ending, the Global Pound Conference platform changed to become the Global Pound Conversation, a blog and research series covering changes and developments in mediation and alternative dispute resolution around the world (https://www.globalpound.org).

Expanded Use of Neutral Third-Parties

Organizations – especially organizations wishing to procure goods and services – have long used outside third-parties such as advisory, consulting and legal service providers to help them select and source suppliers. In the early 1990’s researchers began to explore the concept of using third-parties as “bridge builders” in relationships (Obstfeld, 2005; Long Lingo and Mahoney, 2010). In 1991 L.D. Brown noted third-parties play the role of “central actor among diverse constituencies” who can be an effective conduit for “ideas and innovations, a source of information, a broker of resources, a negotiator of deals, a conceptualizer of strategies, and a mediator of conflict (Brown, 1991).”

Neutral third-parties have been shown to provide value in a variety of capacities, including:

- Research into trust theory suggests that people engage in self-monitoring and demonstrate more trusting behaviors in a relationship when there is a shared link to a third-party (Zucker, 1986)

Unpacking the Standing Neutral

- James Coleman (1990) argued that when a mutual third-party is connecting two parties, the neutral third-party could exert sanctions that will restrain the parties from behaving opportunistically towards each other
- "Bridging organizations" whose role is to facilitate multi-sector partnerships (Westley and Vredenburg, 1991)
- "Boundary spanners" in networks (Burt, 1992)
- "Go-betweeners" in alliances (Noteboom, 2004)
- "Intermediaries" for improving innovation (Howells, 2006)
- Building and repairing trust (Noteboom, 2004; McEvily and Zaheer, 2004; Mesquita, 2007)
- Preventing opportunistic behavior (Coleman, 1990) and reducing the negative effect of power disparity (Noteboom, 2004)
- Henry Adobor and Ronald S. McMullen (2014) found the use of a credible neutral third-party exerts indirect influence by inspiring self-monitoring with no direct sanctions
- Gillian Hadfield (2017) advocates for a shift to third-party regulation to create lower cost approaches to "ensure that not only poor-country suppliers, but also the global corporations that buy from them, are bound to rules"
  Most recently, research in a *Journal of Purchasing and Supply Management* article titled

"Strategic Purchasing and Supplier Partnerships – The Role of a Third-party Organization" shows a third-party can play a "significant and positive role in the development of interfirm relationships (Adobor and McMullen, 2014)."

Advent of the Standing Neutral

As noted previously, the first Dispute Review Board (DRB) was invented in 1975. The DRB was a trusted three-party panel of independent expert advisors chosen by contracting parties to be immediately available to help resolve disputes that arise between them during their contract relationship. By 1991 the process had been used successfully on over 100 projects requiring expertise in only a single technology, such as tunnels (geotechnical engineering), dams (civil engineering), other massive civil engineering projects, and a few commercial projects. By that time the DRB and was recognized as a superior process for keeping the peace on a construction project. The first use of the term "Standing Neutral" to characterize a Dispute Review Board, and its likely applicability to multi-disciplinary types of construction, appears to have been in a 1991 CPR Publication "Preventing and Resolving Construction Disputes (Groton, 1991)."

Unlike a neutral used on an *ad hoc* basis for dispute resolution in mediation or arbitration, a Standing Neutral is a readily-available "fast response" technique, designed to prevent any issues from escalating into adversarial disputes that might otherwise go to mediation, arbitration or litigation. A key feature is that the neutral is "standing." - meaning that it is integrated into the parties' continuing governance structure. Another key concept is that the standing neutral supports the relationship itself and both parties equally; the goal is to ensure the success of the relationship.

The classic Standing Neutral process can best be described as "a quick, informal, flexible, adaptable, non-adversarial, neutral, fact-based, expert, preferably nonbinding, process for proactively achieving the earliest possible solution to problems and potential disputes." The modern Standing Neutral process takes it a step further and is a mechanism for avoiding problems and disputes in the first place. There is little formal research into the use of Standing Neutrals outside of the construction industry. This is a major reason this article is so imperative.

DEVELOPING A STANDING NEUTRAL

The Role of the Classic Standing Neutral

The role of a "classic" Standing Neutral is to serve as a "real time" dispute-resolver throughout a relationship (Groton, 2009). Because they are "standing" they can act immediately to resolve any potential or actual disputes which the parties cannot resolve themselves. There are several variations of a classic Standing Neutral, but almost all involve these common steps:

1. At the outset of their relationship, parties select one or three persons in whom they have trust and confidence to serve as their dispute resolver (the Standing Neutral) throughout their relationship.

2. A single Standing Neutral should always be entirely independent. In most cases where there is a multi-member Standing Neutral, each party nominates one member, and the two nominated neutrals will select a third member; in such cases, it is typically required that every panel member be acceptable to both parties, and that all panel members be independent and impartial, without any special allegiance to the nominating party.

3. Depending on the wishes of the parties, the Standing Neutral is given authority to act on issues and disputes by rendering either a nonbinding evaluation or recommendation, or a binding decision. If the Standing Neutral is empowered to only make a recommendation, either party may challenge the Standing Neutral's recommendation. However, the recommendation will typically be admissible as evidence in any subsequent arbitration or litigation.

4. The Standing Neutral is briefed by the parties regarding the nature, scope and purpose of the relationship or venture. It is usually equipped with a basic set of contract materials and supporting documents.

5. The Standing Neutral is usually part of ongoing governance, to be available on short notice and meet regularly with the parties for a basic review of the progress of the relationship, even if there are no issues. Sometimes the Standing Neutral is merely available on an ad hoc basis, with the contracting parties calling in the Standing Neutral, whenever necessary, for advice, a prompt recommendation, or decision.

6. The parties equally absorb the cost and expenses of the Standing Neutral.

The role of a Standing Neutral has also been referred to as a "Dispute Review Board", "Referee", or "Wise Person".

Three Critical Elements of the Standing Neutral Process

There are three critical elements essential to the success of the Standing Neutral technique:

- Early mutual selection
- Continuous involvement by the neutral

- Prompt action on any issues.

EARLY MUTUAL SELECTION

Using a Standing Neutral begins when the parties mutually agree and designate a single neutral (or a board of three neutrals such as a Dispute Board in the construction industry). The parties should mutually select a Standing Neutral where they have high confidence in the neutral's integrity and expertise. As such, a Standing Neutral is typically an expert in the subject industry the parties are involved in (e.g., construction, facilities management, IT services).

The Standing Neutral should be jointly selected by the parties *early* in the relationship. If the Standing Neutral will play a role as a deal architect, they should be selected *prior* to the parties starting their contracting process. If the Standing Neutral will be used primarily as a cost-effective issue resolution process as part of ongoing governance, the Standing Neutral should be selected *during* the contracting process before the contract is signed. This allows for the Standing Neutral to be embedded as part of the ongoing governance mechanisms.

By establishing a Standing Neutral from the inception of the relationship, the Standing Neutral becomes part of the team and helps to create a collaborative atmosphere. As such, Standing Neutrals are often thought of in a different fashion, which creates an atmosphere far more collaborative than when selecting a mediator or arbitrator after a dispute arises. Many view a Standing Neutral as a "mutual friend," "referee," or "sensible sounding board" because their advice is respected and accepted more readily than if the parties brought in a third-party stranger (mediator or arbitrator) after there is a formal dispute.

Pre-selecting a Standing Neutral at the onset of a contract avoids many problems associated with adversarial jockeying and delays associated with trying to find a suitable mediator or arbitrator after controversy arises. The ready availability of the Standing Neutral and the neutral's familiarity with the relationship make it possible to obtain a prompt resolution of any disputes.

CONTINUOUS INVOLVEMENT

Once the Standing Neutral is selected, he or she is briefed on the relationship and furnished with the basic documents describing the relationship. The role of a Standing Neutral will vary based their entry point into the relationship. For example, the University of Tennessee's popular Vested outsourcing methodology for developing highly collaborative win-win outsourcing relationships embeds a neutral third-party "deal architect" as a coach as part of the contract development (Vitasek and Manrodt, 2012). The Standing Neutral provides an objective view on facts and issues and helps the parties ensure they get to a fair and balanced contract.

Organizations that embed a Standing Neutral as part of ongoing governance (e.g., such as in the construction industry's Dispute Review Board) will rely on the Standing Neutral to help the parties immediately address and resolve issues and concerns that arise in the relationship and to prevent issues from escalating into full-fledged disputes. This early and swift involvement ensures issues are resolved while they are small, preventing the need for more costly mediation, arbitration or litigation.

One of the key differences between a Standing Neutral and a mediator or arbitrator is the fact the Standing Neutral has ongoing involvement with the parties during the life of the contract (or project, as in the construction industry). The parties routinely provide the Standing Neutral with periodic progress reports as the relationship progresses and, when possible, invite the Standing Neutral to meet occasionally with the parties absent any immediate dispute. For example, in construction projects Dispute Boards are often part of the project administration. Likewise, in an outsourcing relationship a Standing Neutral can be embedded into formal governance mechanisms such as Quarterly Business Reviews (Groton and Dettman, 2011).

Because a Standing Neutral has more of a "hands-on" approach, they can almost always earn trust quickly as being fair and impartial. In addition, the continuous involvement of the Standing Neutral enables them to maintain a feel for the dynamics and progress of the relationship. Also, they can coach each party about the potential opportunistic behaviors that can easily start a downward spiral of tit-for-tat negative actions.

A key benefit of having a Standing Neutral embedded in the relationship is that it significantly increases the speed with which he/she can offer advice and render decisions if needed. In addition, the Standing Neutral will hear every dispute that occurs during the history of the relationship, which promotes more candid discussions. This enables the Standing Neutral to shift the focus from that of a “judge” to one of a “coach.”

Last, the ongoing nature of the relationship with the Standing Neutral becomes a powerful inherent incentive for the parties to “self-monitor” behaviors and avoid opportunism and shading (Hart and Moore, 2008) behavior much the way a referee works in a sport to curb bad behaviors. Thus, the Standing Neutral can influence, during the contract period, positive behaviors that improves contract performance. The concept of shading was developed by Nobel Laureate Oliver Hart. Shading is not opportunistic behavior, but rather retaliating behavior in which a party stops cooperating, ceases to be proactive, or makes countermoves because of disappointment. Shading happens when a party does not get the outcome they expect from the deal and feels the other party is to blame for it or does not act reasonably by helping to mitigate the losses. In short – each party’s action leads to the other party’s reaction, creating the negative vicious cycle. One disappointment leads to another and the cycle begins. The concept of shading makes sense, especially with complex deals. In complex deals, a contract will always be incomplete with gaps, errors or omissions opening the door for shading behavior after the contract is signed. Traditional contracts simply do not contain mechanisms to avoid disappointments.

REAL-TIME AND PROMPT ACTION ON ISSUES/CONCERNS/DISPUTES

A key objective of a Standing Neutral process is to preserve cooperative relationships between the contracting parties. The classic Standing Neutral emphasizes “keeping the peace” in a relationship while modern Standing Neutrals focus on a more proactive continual alignment of interests. A good Standing Neutral process is a “fast response/dose of reality” technique emphasizing “real-time” resolution which can be deployed during the contracting process and after the contract is signed (Groton, 2009).

The Standing Neutral is expected to be available on relatively short notice to consult with the parties and to discuss issues while misalignment and problems are still new and likely still small. The Standing Neutral has a remarkable ability to help the parties resolve any misalignment because they are a trusted “part of the team.” The Standing Neutral reviews an issue while it is in the earliest stage and helps the parties identify ways forward in an informal capacity before issues become disputes. The Standing Neutral's early involvement creates valuable opportunities for the parties to avoid disputes through proactive communication. In addition, the participation of lawyers as advocates for parties is discouraged to preserve the informality of the process and to help keep the process as non-adversarial as possible.[*] This offers a significant advantage over traditional mediation, arbitration or reconciliation techniques.

In most cases, if the parties cannot reach a resolution, the Standing Neutral will render an impartial recommendation (not a compromise proposal) when issues arise (Vorster, 1993; Hafer, 2010). If the Standing Neutral is called on to make a recommendation, recommendations are often only regarding matters of entitlement, leaving discussion of amounts up to the parties after they have received the recommendation. Typically, the recommendations of the Standing Neutral are non-binding and parties can opt for a more formal dispute resolution process such as arbitration or litigation if the Standing Neutral's recommendation is not accepted. However, sometimes the parties give the Standing Neutral the authority to act as an arbitrator to make binding decisions. The downside to asking for a binding decision is that this will likely encourage the participation of lawyers serving in an adversarial capacity, changing somewhat the nature of the process.

Experience has shown that when an issue is referred to the Standing Neutral, the Neutral's decisions have generally been accepted by both parties with no attempt to seek relief from any other tribunal. This result is enhanced where there is a contract stipulation stating that the event of any subsequent arbitration or litigation, the decisions of the Standing Neutral will be admissible in evidence in formal arbitration or litigation.

Why the Standing Neutral Process Works So Well

When parties combine the three elements above into a Standing Neutral process, they are in essence establishing the "rules" of how they will use the Standing Neutral to prevent or resolve issues early. A well-designed Standing Neutral embeds their customized rules as a foundational component of the parties' ongoing governance (Groton and Dettman, 2011).

Standing Neutrals have had a remarkable record of success wherever they have been used. In the large majority of cases, the parties never look to the Standing Neutral to make any recommendations or decisions. And in the small minority of cases where the Standing Neutral actually makes a recommendation, 95% of the recommendations are accepted by the parties without resort to mediation, arbitration or litigation (Mathews et al., 1996).

The establishment of a Standing Neutral—which appears at first to be merely an efficient technique for resolving disputes—creates a dynamic situation in which the participants in the business enterprise change their relationship and their attitudes toward each other. The changes usually are an evolution, rather than a conscious effort. For example, at first it is common for contracting parties to feel they are simply choosing an expert neutral for resolving conflicts between them promptly. However, as the Standing Neutral interacts with the parties during ongoing governance forums, the parties develop a greater sense of confidence in the Standing Neutral's ability to quickly alleviate friction in the relationship. When this happens, the parties shift their view of the Standing Neutral from "dispute resolver" to a view of "mutual friend" or a "sensible sounding board."

[*] *Lawyers are often Standing Neutrals. However, when acting in the capacity of a Standing Neutral they do not formally represent either party in a formal legal capacity, but rather as a neutral advisory role*

Unpacking the Standing Neutral

Simply put, the mere act of appointing a Standing Neutral can be like a magic bullet for reducing or even eliminating friction between parties to a contract. Research supporting this dates to 1933 when Elton Mayo (Mayo 1933; Shafritz et al., 2016) conducted research into the “Hawthorne Effect,” which states the mere act of watching can affect behaviors. Since then several researchers have shown the impact of using outsiders. For example, Adobor and McMullen (2014) found “the sheer presence of a third-party fosters ‘self-monitoring’ of behaviors” and Dan Ariely has shown that the presence of others causes people to behave more honestly and reign in unethical behavior such as cheating (Ayal et al., 2015). These effects are amplified when the third-party observer is knowledgeable in the subject matter of the agreement and in the nature of the agreement. Table 1 illustrates how each of the examples is being used in practice.

TABLE 1. The timeline for adding a Standing Neutral to a contract

<table>
<tr><th>Pre-Contract Signing</th><th colspan="5">Post-Contract Signing</th></tr>
<tr><th colspan="2">Problem Prevention Techniques</th><th>Problem Solving Techniques</th><th>De-Escalation Techniques</th><th>Dispute Control Techniques</th><th>Formal Dispute Resolution Methods</th></tr>
<tr><td colspan="6">Branding & Licensing Example</td></tr>
<tr><td></td><td></td><td></td><td colspan="2">Franchise Wise Persons Committee</td><td></td></tr>
<tr><td></td><td></td><td>Outsource Agreement Embedded Governance</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td><td colspan="3">Real Estate Development Standing Arbitrators</td></tr>
</table>

Unpacking the Standing Neutral

<table>
<tr><td>Labor
Services
Deal
Architect</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td colspan="3">Non-Profit NGO Wise Persons Committee</td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Toyota
Standing
Arbitrator</td></tr>
<tr><td></td><td></td><td colspan="2">South Korea Ombudsman
Office</td><td></td><td></td></tr>
<tr><td></td><td></td><td colspan="2">Outside Director Role</td><td></td><td></td></tr>
<tr><td colspan="3">Telia/Veolia Deal Architect</td><td></td><td></td><td></td></tr>
</table>

[Detailed information and reference citation removed for blind peer review]

While the cost of mediation in and of itself is not expensive the largest cost is one of lost time, lost trust, and legal and consulting fees that add up over the protracted period. Figure 2 illustrates the relative cost differences of different methods of dispute resolution.

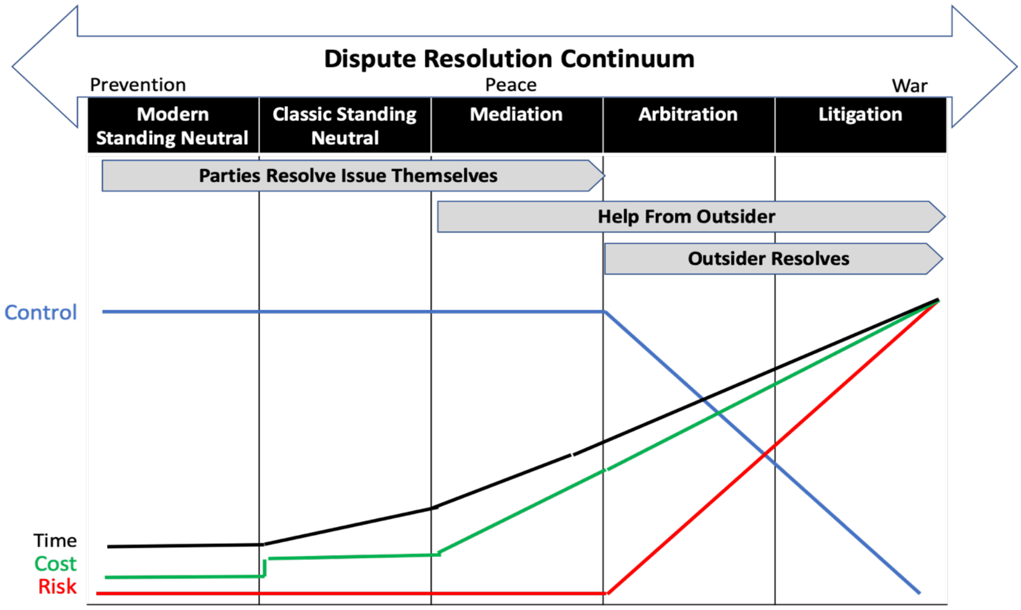
FIGURE 2. Relative Transaction Costs of Different Methods of Dispute Resolution

Judicial Proceedings	$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
Arbitration	$$$$$$$$$$$$$$$$$$$$$$$
Mini-Trial	$$$$$$$$$$$$$
Mediation	$$$$$$
Expert Advisory Opinion	$$$
Standing Neutral	$

Unpacking the Standing Neutral

Figure 3 expands on the notion of the continuum of cost, risk and control of dispute resolution originally developed by Kelsey to include a Standing Neutral and the aspect of time.

FIGURE 3. Dispute Resolution Continuum



CONCLUSION

While Louis M. Brown is credited as the founding father of “preventive law,” his early work inspired a growing cadre of followers who have researched and expanded on every facet of the concept of preventing disputes. Today there is a clear and unmistakable evolving trend toward incorporating proactive approaches for preventing and managing disputes into all business relationships. This recent trend is aptly termed “the Prevention Movement.”

The use of a Standing Neutral in business relationships – especially a modern Standing Neutral who focuses helping the parties stay in continual alignment - proves the adage “An ounce of prevention is worth a pound of cure.”

While there are skeptics, the Prevention Movement is taking hold as evidenced at the 2017-2018 Global Pound Conferences held worldwide. During the conference major stakeholders in the dispute resolution field (users of dispute resolution services, their advisors and lawyers, providers of both adjudicatory and non-adjudicatory services, and the researchers and educators who influence the users of dispute resolution services) revealed the following consensus:

- Dispute resolution should be conceived and practiced earlier in the trajectory of risks that can develop into conflict, escalating from differences of opinion to arguments, aggression, and finally disputes that have to be dealt with through formal dispute resolution efforts.
- Pre-dispute or pre-escalation techniques are the most promising and valuable methods for improving the future of dispute resolution and should prevent disputes
- Where possible, risks should be understood and addressed in advance so problems never arise.
- Where efforts to prevent problems fail, steps should be initiated to de-escalated, contain, or provide "real time" resolution of conflicts so the costs, hostilities and delays of formal dispute resolution can be avoided.

The conclusions from the Global Pound Conferences demonstrate that the Prevention Movement is no longer just an aspiration of a few visionaries – but one that is seen as needed in today's modern economy.

This article provides a framework for use of a Standing Neutral and how a Standing Neutral can be incorporated into all facets of a business relationships – ranging from pre-contract signing all the way through formal dispute resolution techniques that are much more effective than traditional mediation and arbitration.

REFERENCES

Adobor, H. and McMullen, R.S. (2014). Strategic purchasing and supplier partnerships—The role of a third-party organization. *Journal of Purchasing and Supply Management*. December 2014, pp. 263-272. Available: https://www.sciencedirect.com/science/article/pii/S1478409214000338

Ayal, S. Gino, F., Barkan, R., and Ariely, D. (2015). Three principles to REVISE people's unethical behavior. *Perspectives on Psychological Science*. 10(6): 738-741. Available: https://www.researchgate.net/publication/284184987_Three_principles_to_REVISE_people's_unethical_behavior

Brestoff, N.E. (2012). Data Lawyers and Preventive Law. *Legaltech News*. October 2012. Available: https://intraspexion.com/data-lawyers-and-preventive-law-1

Brestoff, N.E. and Inmon, W.H. (2015). *Preventing Litigation: An Early Warning System to Get Big Value out of Big Data*. New York: Business Expert Press.

Brown, L.D. (1991). Bridging Organizations and Sustainable Development. *Human Relations*. 44 (8), 807-831, 1991.

Brown, L. M. (1950). *Manual of Preventive Law*. Prentice-Hall.

Burt, R.S. (1992). *Structural Holes: The Social Structure of Competition*. New York: Harvard University Press.

Carver, T.B. and Vondra, A.A. (1994). Alternative Dispute Resolution: Why It Doesn't Work and Why It Does. *Harvard Business Review*. May-June 1994. Available: https://hbr.org/1994/05/alternative-dispute-resolution-why-it-doesnt-work-and-why-it-does

Center for Public Resources (1991). Preventing and Resolving Construction Disputes. *Center for Public Resources*. Dec. 1991. Available: https://doi.org/10.1002/alt.3810091208

Coleman, J.S. (1990). *Foundations of Social Theory*. Cambridge, MA: The Belknap Press of Harvard University Press.

Costantino, C.A. and Sickles Merchant, C. (1996). *Designing Conflict Management Systems: A Guide to Creating Productive and Healthy Organizations*. San Francisco: Jossey-Bass.

Construction Industry Institute (1986). Incentives in Construction Contracts. CII Source Document SD-8. *Construction Industry Institute*. Available at: https://www.construction-institute.org/resources/knowledgebase/knowledge-areas/procurement-contracts/topics/rt-005/pubs/sd-8

Construction Industry Institute (1988). Contract Risk Allocation and Cost Effectiveness. *Construction Industry Institute*. Available: https://www.construction-institute.org/resources/knowledgebase/knowledge-areas/procurement-contracts/topics/rt-005/pubs/rs5-3

International Institute for Conflict Prevention & Resolution (2010). Reducing Disputes Through Wise Prevention Processes in Business Agreements. CPR Prevention Practice Materials. November 2010. Available: https://www.cpradr.org/news-publications/articles/2010-11-23-cpr-prevention-practice-materials

Diekmann, J.E., Girard, M.J., Abdul-Hadi, N (1994). DPI-Disputes Potential Index: A Study into the Predictability of Disputes. *Construction Industry Institute*. Austin, Texas.

Dunnewold, M. (2009). What Every Law Student Should Know. *Student Lawyer*. October 2009. 38(2). Available: https://www.americanbar.org/content/dam/aba/migrated/2011_build/dispute_resolution/what_every_law_student_should_know.pdf

Groton, J.P. (1990). Dispute Resolution Devices for the Construction Industry: An Overview. *The Punch List*. 13(3-4).

Groton, J.P. (1991). Jobsite Dispute Resolution. *Preventing and Resolving Construction Disputes* (Chapter 6). CPR Institute. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/alt.3810091208 and https://www.cpradr.org/resource-center/toolkits/construction-briefing-dispute-resolution-boards-other/_res/id=Attachments/index=0/CPR-Dispute-Review-Boards-and-Other-Standing-Neutrals-Constructiontitle.pdf

Groton, J.P. (2009). The Standing Neutral: A 'Real Time' Resolution

Procedure that also Can Prevent Disputes. *Alternatives to the High Cost of Litigation* 27(11): 177-185.

Groton, J.P. and Dettman, K.L. (2011). How and Why the Standing Neutral Dispute Prevention Technique Can Be Applied. *Alternatives to the High Cost of Litigation* 29(10): 177-192.

Hadfield, G.K. (2017). *Rules for the Flat World*. New York: Oxford University Press.

Hafer, R. (2010). Dispute Review Boards and Other Standing Neutrals: Achieving 'Real Time' Resolution and Prevention Disputes. CPR Dispute Prevention Briefing: Construction. Available: https://www.cpradr.org/resource-center/toolkits/construction-briefing-dispute-resolution-boards-other/_res/id=Attachments/index=0/CPR-Dispute-Review-Boards-and-Other-Standing-Neutrals-Constructiontitle.pdf

Hart, O. and Moore, J. (2008). Contracts as Reference Points. *The Quarterly Journal of Economics*. CXXIII(1). Available: https://scholar.harvard.edu/files/hart/files/contractsasreferencepointsqje.pdf

Howells, J. (2006). Intermediation and the role of intermediaries in innovation. *Research Policy*. 5. pp. 715–728.

Legal Information Institute (2017). Alternative Dispute Resolution. Cornell Law School. Available: https://www.law.cornell.edu/wex/alternative_dispute_resolution

Long Lingo, E. and Mahoney S. (2010). Nexus Work: Brokerage on Creative Projects. *Administrative Science Quarterly*. March 2010. Available: https://journals.sagepub.com/doi/abs/10.2189/asqu.2010.55.1.47?related-urls=yes&legid=spasq%3B55%2F1%2F47

Mathews, A.A., Smith, R.J., Sperry, P.E., Matyas R.M. (1996). *Construction Dispute Review Board Manual*. New York: McGraw-Hill.

Mayo, E. (1933). The Hawthorne Experiment. Western Electric Company.

McEvily, B. and Zaheer, A. (2004). Architects of trust: the role of network facilitators in geographical clusters. *Trust and Distrust in Organizations*. New York: Russell Sage Foundation. pp. 89–213.

McManus, M. and Silverstein, B. (2011). Brief History of Alternative Dispute Resolution in the United States. *Cadmus*. 1(3). Available: https://www.cadmusjournal.org/article/issue-3/brief-history-alternative-dispute-resolution-united-states

Mesquita, L.F. (2007). Starting over when the bickering ends: rebuilding aggregate trust among clustered firms through facilitators. *Academy of Management Review*. 31(1), 72–91.

Noteboom, B. (2004). *Interfirm Collaboration, Learning and Networks: An Integrated Approach.* London: Routledge.

Unpacking the Standing Neutral

Obstfeld, D. (2005). Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. *Administrative Science Quarterly*. Available: https://journals.sagepub.com/doi/10.2189/asqu.2005.50.1.100

Shafritz, J., Ott, S., and Jang, Y.S. (2016). *Classics of Organization Theory*, 8th Edition, Cengage.
Vitasek, K., & Manrodt, K. (2012). Vested outsourcing: a flexible framework for collaborative outsourcing. *Strategic Outsourcing: An International Journal*, 5(1), 4-14. Available: https://www.researchgate.net/profile/Kate_Vitasek/publication/263240305_Vested_outsourcing_A_flexible_framework_for_collaborative_outsourcing/links/54b2fab30cf2318f0f953bc3.pdf.

Vorster, M.C. (1993). Alternative Dispute Resolution in Construction with Emphasis on Dispute Review Boards. CII Source Document SD-95. *Construction Industry Institute*. Available: https://www.construction-institute.org/resources/knowledgebase/best-practices/disputes-prevention-resolution/topics/rt-023/pubs/sd-95

Westley, F. and Vredenburg, H. (1991). Strategic bridging: the collaboration between environmentalists and business in the marketing of green products. *The Journal of Applied Behavioral Science*. 27(1), pp. 65–90.

Williams, O.E. (2008). Outsourcing: Transaction Cost Economics and Supply Chain Management. *Journal of Supply Chain Management* 44(2). Available: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1745-493X.2008.00051.x

Wolaver, E.S. (1934). The Historical Background of Commercial Arbitration. University of Pennsylvania Law Review. December 1934. pp.132-146. Available: https://scholarship.law.upenn.edu/penn_law_review/vol83/iss2/2

Zucker, L.G. (1986). Production of trust: institutional sources of economic structure. *Research in Organizational Behavior*. p. 53–112. Greenwich, CT: JAI Press.

DECISION SCIENCES INSTITUTE

Investigating Student Adoption of AI Teaching Assistants from Innovation Diffusion Theory and The Anthropomorphism Perspectives

My-Trinh Bui
VNU International School - Vietnam National University, Hanoi, Vietnam
Email: trinhbm@isvnu.vn

Ngoc - Ai Jessie Pham
Graduate Institute of Management - Minghsin University of Science and Technology, Taiwan
Email: phamngocai@gmail.com

Don Jyh-Fu Jeng
Graduate Institute of Technology, Innovation and Intellectual Property Management National Chengchi University, Taiwan
Email: jeng@nccu.edu.tw

ABSTRACT

Virtual-agent instructors, also known as artificial intelligence (AI) teaching assistants, were created in response to a rise in demand for online education. AI teaching assistants have gradually been incorporated into a small number of courses. However, few studies have been conducted to acknowledge how students would react to AI-assisted instruction, especially their venturesomeness, functional and social perception of AI teaching assistants. The current study evaluated students' standpoints of AI teaching assistants in higher education. The perceived functionality and sociality of AI teaching assistants are crucial to understanding the ultimate adoptionof AI teaching assistant-based education.
KEYWORDS: Venturesomeness, functional instruction, parasocial relationship, adoption

INTRODUCTION

With the increasing demand of smart technologies and the advances in Digital Life, the artificial intelligent virtual assistant market is forecasted to flourish from USD2.48 billion in 2017 to USD17.72 billion by 2023 (ResearchandMarkets, 2017). That is a promising figure when the world is practically running on computerized reasoning, for example, Siri oversees our schedules, Facebook recommends our companions and PCs exchange our stocks. Each field has profited by progress in man-made consciousness, from the military to medication to assembling. Although online education has rapidly evolved from traditional education over the last decade, the new technology era expands further. The arrival of AI teaching assistants signals the start of a new trend of education, with nonhuman agents serving as tutors, assistants, consultants, and/or professors, so-called “machine teachers.”
While machines will not be able to entirely replace human teachers, they do have the ability to play a variety of roles in education. Machines are increasingly being advanced to teach and study by collaboration, and to be focused on natural instruction and learning tactics utilized by their human cohorts (Edwards et al., 2016). As a result, we are rapidly entering a modern age in education in which computers become part of the instructor pool. It is necessary to improve education effectiveness, commit to learning efficiency, and enhance the capability of each

student. The difficulties that should be routed to make these enhancements enormously surpass the extent of any single methodology, regardless of whether it is instructive innovation, improved instructor preparation, and better after-school programs.

This situation raises a number of critical concerns. What, for example, are the relationships between the perceived technology venturesomeness of computer-based instructor and student adoption of new educational method, and how do these relationships mediate social and functional teaching to satisfy student studying needs? What is the role of a computerized teacher in classroom management? Before we explore these and other issues, it's important to have a basic understanding of what's the extent, students are willing to consider the notion of computer teachers during their study process. While the technology to build an AI teaching assistant is advanced, few researches have been done to understand how students will respond to the AI teaching assistant.

In the online education context, particularly during the episode of COVID 19 and later on, AI innovation will without a doubt be eagerly invited and broadly utilized and generally utilized as an application. Given that AI in education is a relatively recent concept, it is critical to comprehend this phenomenon from the perspective of technology users, especially students. To contribute to this body of knowledge, the current study investigates how students' adoption of AI-based educational mechanism is linked to their experiences of AI teaching assistants. The study focuses on student’s venturesomeness to increase their parasocial relationship and functional instruction with AI through anthropomorphism theory.

LITERATURE REVIEW

AI Technology in Education

In the education context, there has been a growth in the use of smart technology. Smart technology such as artificial intelligence virtual assistants (AI VA) are self-contained devices that obey social behavior standards and communicate with humans in a variety of environments (Gockley et al., 2007). Despite the fact that AI VAs were first used in schools in the 1980s, they are becoming now commonplace (Johal et al., 2018). In educational environments, smart technologies may play the part of an instructor, mentor, peer, or even a healthcare helper (Belpaeme et al., 2018). More specifically, an AI-powered virtual teaching assistant is a program designed to improve a student’s learning experience and to take some load off the faculty’s shoulders. With machine learning capabilities, these programs don’t just provide answers to common questions, they can analyze the learning process and make suggestions to optimize the approach to education.

Chou et al. (2011) proposed that virtual tutors for each student are omnipresent help that incorporates individual demonstrating, social reproduction, and information portrayal. Besides, it also addresses 21st-century abilities assisting students with self-heading, self-appraisal, collaboration. Throughout the most recent decade, the utilization of AI has tended to handle difficulties, including language handling, thinking, arranging, and psychological display (Mishra et al., 2020, Tuomi, 2018). Known as Intelligent Tutor Systems which is a one-by-one teaching system, PC programming is capable of tracking the "psychological strides" of the student during critical thinking errands to analyze confusions and gauge the student's comprehension of the area. Intelligence Tutor Systems additionally can give opportune direction, criticism, and clarifications to the student and can advance profitable learning practices, like self-guideline, self-checking, and self-clarification.

Besides, Intelligent Tutor Systems can likewise recommend learning exercises at the degree of trouble and with the substance generally proper for the student (Vanlehn, 2011). These frameworks are additionally ready to emulate the advantages of balanced coaching, and a portion of these frameworks outflank undeveloped mentors in explicit themes and can move toward the adequacy of master guides (Vanlehn, 2011). In another learning application, man-made reasoning can help coordinate and incorporate substance to help content conveyance. Known as profound learning frameworks, innovation can peruse, compose and imitate human conduct. For instance, Dr. Scott R. Parfitt's Content Technologies, Inc. (CTI) empowers instructors to gather custom course books. Teachers import a prospectus and CTI's motor populates a coursebook with the center substance.

As with disputes about the importance of offering courses entirely online, the concept of using robots in teaching has sparked differing viewpoints among educational researchers (Javaheri, 2019). Some contend that students need a human teacher, despite the fact that student's information recall is dramatically higher when provided by a robot and smart technology (Li et al., 2016). Previous investigations show that the faculty hold a strong degree of reluctance to embrace different forms of online teaching, due to fear of change, concerns about the reliability of technology, skepticism about student outcomes in online learning environments, workload issues, and other factors (Sorokova, 2020). Other factors such as higher price, teacher instruction, and applicability are also among the other arguments (Johal et al., 2018).

Despite the above-mentioned disagreements, a large body of research has shown that smart AI technology has a beneficial impact in pedagogical settings. AI robots were considered to be the worthwhile knowledge provider for student learning by a group of scholars (Edwards et al., 2016, Park et al., 2011). Park and colleagues (2011) found a robot mentor who offered constructive feedback was viewed as desirable and appropriate when teaching participants on a specific subject. Kim et al. (2020) also recognized that autonomous systems would replace human in the near future. In a related vein, it is believed that students have enhanced their performance with the support of AI-Driven Virtual Teaching Assistants (AI TAs) by actively engaging in a specific learning environment and synchronously interacting with peers and teachers (DoubleRobotics, 2017). Overall, these findings back up a study of evidence that shows that AI TA in educational environments improves student learning (Belpaeme et al., 2018).

Although AI Teaching Assistants have the potential to improve student learning outcomes, there are a few factors that should be taken into account if AI TA is to be used more successfully as a pedagogical medium. Castellano et al. (2013) suggested that empathy and motivation are key successful factors for student machine-based learning. Li et al. (2016) discovered that among various robotic technologies, a physically present AI TA would send more convincing messages and attract more attention than a virtual agent (e.g., a remotely present robot on a screen). Furthermore, Li et al. (2016) discovered that while a video of a human teacher and a machine counterpart have comparable effects on students' information recall, a video of a robot teacher has poorer effects on recall results.

Conclusively, most research on robotics in education has found positive approaches to facilitate successful learning interactions The beneficial benefits of using AI TAs, or machines in a broader sense, can vary depend on a number of variables, including the forms of instruments and the teaching materials. However, given that robots can be useful in promoting successful teaching, especially in periods when health and safety should be the priorities such as during Covid-19 pandemic, there appears to be a case for considering the concept of computer teachers in education.

Innovation Diffusion Theory: AI Venturesomeness and Meet The Expectation Of Social, Functional Relationship

Explanations and forecasts regarding market acceptance of emerging technologies and the diffusion process have aroused academic and industrial interest, and several similar modeling models have been suggested. Among these, Innovation Diffusion Theory (IDT) and Technology Acceptance Model (TAM) are the most prominent. Rogers (1995) proposed IDT by methodically assembling a variety of experiments on innovation acceptance. According to the theory, innovation is defined as a new idea, practice, or object which is perceived by an individual or other unit of adoption. Rogers (1995) argued that diffusion is the process by which an innovation is communicated over time among the participants in a social system and he classified consumers into five categories according to their innovativeness. IDT has been often used to classify innovation acceptance, dissemination processes, and influence variables (Lee et al., 2011).

IDT has been employed in empirical research to validate the impact of interpretation on innovation acceptance for five innovation characteristics. The relative benefit, compatibility, trialability, and observability of an invention are well known in IDT, and a low understanding of its difficulty is consistent with its rapid dissemination (Kleijnen et al., 2009). The perception of each of these characteristics is assumed to have a strong relationship with the innovativeness of an individual. IDT, therefore defines innovativeness, a psychological trait of users, as another aspect that can justify the introduction of new technologies.

In psychology, innovativeness is described as “venturesomeness,” or variety-seeking and novelty-seeking, and is regarded as the most distinguishing market trait to describe innovation acceptance (Rogers, 2003). Plog (2002) defined venturersomeness as people’’ unique characteristic that drive people to take risk by engaging in a dangerous but potentially rewarding jouneys. By constantly seeking new experiences to enrich their daily lives, they build up their such strong sense of self-confidence so that it allows them successful at whatever they try.

In the tourism context, venturesomeness was defined as being comparable to novelty and variety seeking (Legoherel et al., 2015), sensation seeking (Lepp and Gibson, 2008), or uncertainty avoidance (Karl, 2018) and it had a significant impact on the decision-making process of tourist segments. The higher an individuals' degree of venturesomeness was, the more novel product choices he made. This observation was consistent with previous research that demonstrated the effects of tourists' personality traits on travel choices (Masiero and Qiu, 2018). Individuals who are self-confident and have diverse preference habits, allocentrics, also known as venturers by Plog (2002), may choose more unconventional, more challenging experiences before others do.

An online classroom is a new environment for students. It requires an adaptation process, the length of which depends on each student’s personal traits, especially students’ self-reliance and confidence. Basinska and Dåderman (2018) appears that venturesomeness was essentially related with self- efficacy. According to Bandura (2001), the stronger students perceived self-efficacy, the higher the goal challenges the students set for themselves and the former is their commitment to them. Besides, Bandura (2001) found that students enhanced perceived self-efficacy had a positive relationship with aspirations and adoption of the new challenges and new learning environments. Kim et al. (2017) also agreed that innovativeness plays a major role in explaining of students’ intention to use mobile learning.

Therefore, we have:

Hypothesis 1: Higher level of students' technology venturesomeness will lead to their higher level of adoption of AI TA

AI As Virtual Teaching Assistants Under the Anthropomorphism Aspect

AI Teaching Assistants are among the advances in the digital technologies available to support and empower online education because of their capabilities to assists humans in engaging in affective, cognitive, and behavioral learning in a variety of ways. Hill (2016) in his sharing on singularityhub.com about Jill Watson, a teaching assistant assigned to moderate an online forum for a computer science class. Jill was actually 1 of 9 AI Teaching Assistants assigned to help answer questions about coursework and projects from the 300 students enrolled in the advanced course at Georgia Tech. The interesting aspect of this story is how students associated personality and characteristics to Jill. Jill was believed to be a real person with all capabilities of creativity, critical thinking, communication and collaboration.

Anthropomorphism, according to (Guthrie, 1993), is the practice of imbuing nonhuman agents with actual or imaginary human-like traits, motives, intentions, or feelings. Consumers' tendency to humanize goods and their product reviews are highly influenced by the degree of human-like characteristics applied to the product characteristics or brands (Qiu and Benbasat, 2014). This idea is not far-fetched, since there have been previous efforts to humanize robotic or automated computer companions for the purpose of assisting with basic household chores (Dautenhahn et al., 2005). Furthermore, it has been suggested that building virtual social connections or infusing technologies with anthropomorphic characteristics may improve users' subsequent encounters (e.g., product linkage or comprehension process). Pleasant, sociable, and extroverted robots, for example, earned more favorable perceptions than extreme or introverted robots, according to Goetz and Kiesler (2002). According to (Moreale and Watt, 2003), an anthropomorphized assistant offering helpful guidance for any scenario inside software programs (e.g., Microsoft Office Word) increases users' willingness to complete tasks. Favorable emotions and improved understanding were found in interfaces with more anthropomorphic faces and voices in an experiment of players participating in a desert survival hunt (Burgoon et al., 2000).

Using anthropomorphism often reinforces social ties and fosters a sense of social relationship, which can increase the perceived usefulness of the humanized object (Goudey and Bonnin, 2016, Belk, 2016). Generating stronger social relations can be achieved by increasing the degree to which a good or service is built to have certain human characteristics (Epley et al., 2007), as exemplified by toymakers generating a link between children and toys of human-like features such as eyes, noses, and ears. Nass et al. (1994) have found that adding feminine voices to electronic devices made audiences think of them as frail, while using more masculine voices made listeners think of them as more compelling and influential. These concepts were considered in order to aid personalization attempts in the creation and development of future models of intelligent agents and systems (Xu and Wang, 2006). Newman (2014) also spoke about how Apple's Siri would be an autistic boy's "only buddy forever" and have social help. These examples back up claims that AI VA users would respond differently to anthropomorphized and objectified AI VAs.

AI VAs, which were originally embedded in mobile phones, can now be used for a variety of tasks, such as determining a user's position using GPS/Wi-Fi reception, tracking his or her movements using an accelerometer and gyroscope, and processing speech patterns using chip-level language processing. AI VAs can also monitor the physical environment (e.g., using temperature, light, and humidity sensors) as well as track medical conditions (e.g., blood pressure and heart rate) (Saeb et al., 2015).

In the context of online learning, there is a trend of developers anthropomorphizing AI TA interfaces, incorporating human voices and natural conversations to make them sound “smart,” witty, friendly, and to communicate like a human being to encourage more student interaction. This is because students may exhibit more complex feelings and behaviors when they interact with another human being as compared with objects and things. This notion could be applied to education, where a higher level of perceived AI TA venturesomeness would lead to a higher level of students and machine’’ parasocial-relationship and functional relationship.

Hypothesis 2: The stronger student’s perceived technology venturesomeness will lead to his higher perceived functional instruction towards AI TA

Hypothesis 3: The stronger student's perceived technology venturesomeness, the higher level of parasocial relationship perceived with AI TA
Hypothesis 4: Higher level of student's perceived functional relationship with AI TA will lead to a higher level of adoption of AI TA
Hypothesis 5: Higher level of perceived parasocial relationship with AI TA, the more likely student is to adoption of AI TA

METHODS

AI Participants

Like many countries, Vietnam education has faced the challenges brought about by the pandemic of COVID-19. The sudden closures of schools, colleges and universities have disrupted the flow of learning and teaching activities. For students in big cities like Ha Noi, the capital of Vietnam, online training is the best way to educate students in the context of social distance and isolation. Initially, 500 undergraduates from different majors at 8 major public universities in Vietnam were invited to participate in this research. Firstly, the survey was administered to different groups. Next, in the middle of the survey, an attention search was performed to ensure that participants scrutinize the survey questions. Incomplete survey was removed from the survey. Finally, 428 were selected for further analysis.

Procedure

Sample and data collection are included 3-stage procedure:
A questionnaire was circulated along with an online survey platform approved by a class's instructor (www.google form). The primary researcher contacted undergraduate course instructors and asked if they would be able to engage their students into the study. Following acceptance, a practical excercise was sent to potential participants. One week before the formal survey, all the participants experienced practising AI to assist their writing essay report as a part of the requirements for their in-class assignment. Most of them use Quillbolt and Grammarly for assignments. AI learning entails acquiring and digesting new learning experience as well as developing new studying behavioral models. Self-correction is the process of fine-tuning algorithms to produce the most accurate output.
Students were asked to fill out the survey during their online class because of the COVID-19 pandemic. After accessing the sample, the researchers continued to reaffirm the students' 'consent before their proceeding to complete the survey. Participants' views toward emerging innovations (e.g., Apple's Siri, Amazon's Alexa) were measured at the start of the survey. Students can avoid embarrassment by asking for aid from AI virtual personal assistants in front of their peers. They were guided to read a report in a magazine about an AI teaching assistant in higher education written in Vietnamese language. The article was about an AI teaching assistant invented by a professor and his team in an anonymous university. The students were asked a question regarding their perception of AI TA which they have learned from the magazine. The purpose was to ensure that all participants went over and understood the article as part of their research participation. Confidentiality and anonymity were guaranteed.
The online survey offered a video clip simulation of a virtual AI assistance (Luna AI). The video was embedded using the Google survey system. Participants spent about 5 minutes watching video with high-quality text-to-speech engine female voice, and a human voice (native female English speaker) via Vietnamese transcript (see Figure 3). After viewing the stimulated video, a self-report pretest on simple questions was administered on 30 students to remove ambiguity.

The final questionnaire was conveyed to 500 students. The survey has been conducted for three months across different universities and college. Missing value and incomplete responses were removed. The final responses of 428 students were used to analyze.

RESULTS

Table 1 lists the respondents' demographic characteristics, including gender, study major, and computer usage habit.

Table 1: Demographics (n = 428)			
Measure	Item	Frequency	Percentage
Gender	Female	265	61.9
	Male	163	38.1
E-devices using hours (for study purpose)	<1 hour	89	20.8
	1-3 hour	253	59.1
	3-7 hour	76	17.8
	>7 hour	10	2.3
Major	Social sciences	214	50
	Nature Sciences	214	50

Measure
All the items were 7-point Likert-type scale (e.g., 1 = Strongly Disagree, 7 = Strongly Agree). Technology Venturesomeness: Respondents completed the 3-item venturesomeness adopted from previous studies e.g. 'When I see a new tool, I just often buy it to see what it is like", "I like to try new and different things", 'I enjoy doing new thing"" (Kim et al., 2017, Eysenck and Eysenck, 1978). The concept of parasocial Relationship with AI Technology is 6-item scale adopted from parasocial interaction (Rubin and Perse, 1987). AI functional Instruction was adopted from study Ozkan and Koseler (2009) and (Nass et al., 1994).
And finally, a set of questions employed to assess participants' adoption of AI teaching assistant were adopted from (Choi and Ji, 2015).

Results
Since this study tested a predictive model, partial least squares structural equation modeling (PLS-SEM version 3) was used, which is a preferable approach where the aim of the analysis is hypothesis growth and prediction (Hair, Hult, Ringle, & Sarstedt, 2014; Hair, Risher, Sarstedt, & Ringle, 2019). PLS-SEM has been widely used in the social sciences, including travel and tourism studies (Do Valle & Assake, 2016). An initial assessment of the outer models using SmartPLS 3.2.7 to conduct route analyses for the two destinations revealed that venturesomeness yielded average variance extracted (AVE) values marginally under.50. Then, the models with the reduced three-item venturesomeness construct were reexamined, showing all AVEs greater than .50 and higher than all the squared correlations (see Table 3), indicating appropriate convergent and discriminant validity (Hair et al., 2014).

The measurements had sufficient distributional properties for PLS-SEM, and variation in fraction factors (VIFs) less than 5.0 indicated that no multicollinearity thresholds were violated (Hair et al., 2014). Table 2 indicates outer loadings greater than.70.

Table 2. Descriptive Statistics and PLS-CFA Results (n=428)

	Item	SFL	Cronbach's Alpha		CR	(AVE)	VIF	Mean	SD
AI adoption	AD01	0.932	0.922		0.951	0.866	3.553	5.557	1.503
	AD02	0.948					3.365	5.385	1.542
	AD03	0.911					3.016	5.303	1.634
Functional Instruction	IA01	0.858	0.925		0.943	0.770	2.939	5.119	1.432
	IA02	0.874					3.354	4.915	1.396
	IA03	0.911					3.740	5.124	1.438
	IA04	0.881					3.271	4.952	1.419
	IA05	0.862					2.862	4.858	1.533
Parasocial relationship	PR01	0.892	0.954		0.963	0.813	3.290	5.007	1.475
	PR02	0.905					3.596	4.947	1.446
	PR03	0.924					3.803	5.257	1.447
	PR04	0.909					3.422	5.243	1.489
	PR05	0.897					3.637	5.028	1.424
	PR06	0.881					3.441	5.144	1.453
AI Venturesomeness	TV01	0.777	0.858		0.914	0.782	1.659	4.356	1.681
	TV02	0.948					3.425	5.154	1.580
	TV03	0.918					3.682	5.353	1.497

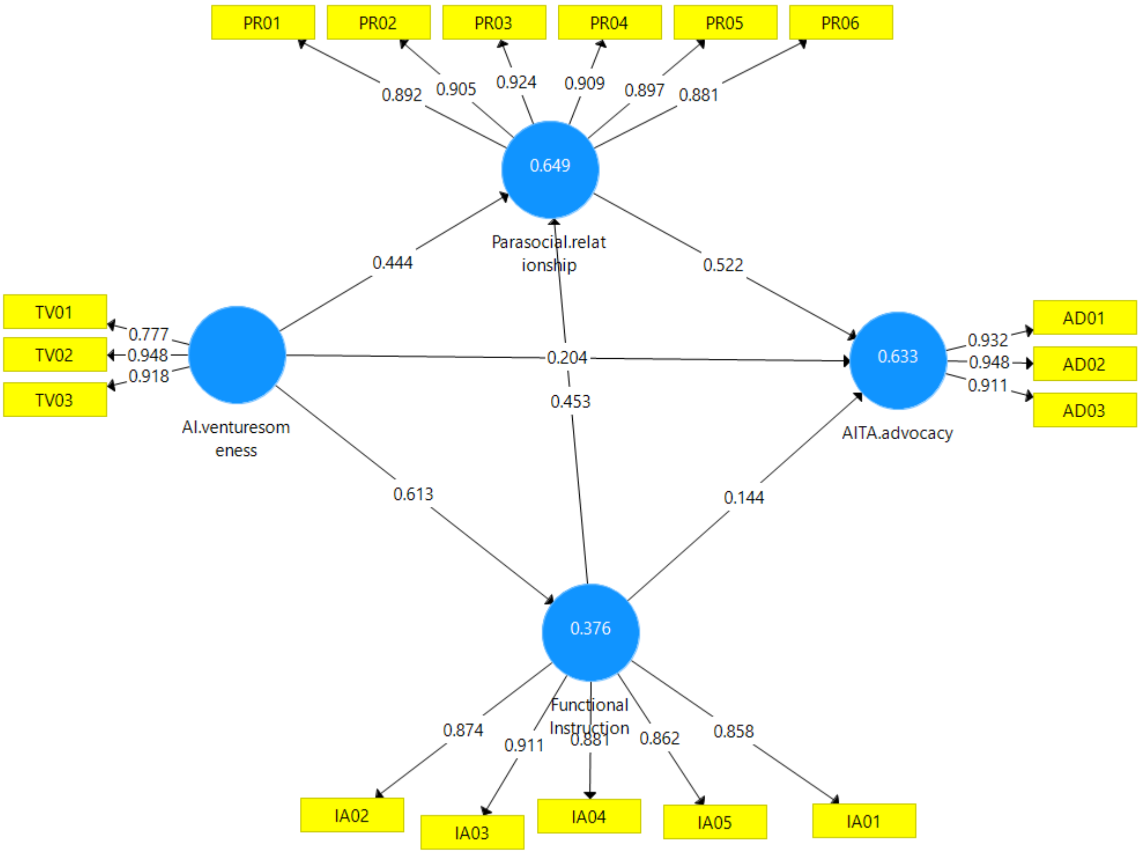
Average Variance Extracted=AVE, SFL=Standardized Factor Loading, VIF, Standard Deviation, Composite Reliability= CR

Table 3: Pearson Correlation

	TV	FI	PS	AD
TV	1			
FI	.509**	1		
PS	.615**	.619**	1	
AD	.559**	.547**	.673**	1

Notes: TV= AI venturesomeness, FI= Functional Instruction, PS= Parasocial Relationship, AD= AI TA adoption

Figure 1. PLS-SEM model



DISCUSSION AND CONCLUSIONS

The current study investigates students' expectations of an AI teaching assistant in higher education under the perspectives of anthropomorphism and diffusion of innovation. Primary results show that students' technology venturesomeness, parasocial relationship and functional relationship with an AI teaching assistant positively predict beneficial adoption of using an AI teaching assistant, which leads to greater motivation to use an AI teaching assistant. Forecasting student behaviors, however, is not a simple process, especially where new technology is deeply concerned. Since AI TAs have become increasingly popular, the literal, mediated personality of this new creature should be included and involved in the research model. Other factors such as the design metaphor, flow of the learning experience, styles of textual, audio, and graphic presentation of information all become the possible effects that encourage parasocial interaction and functional relationship between students and the AI TA based online learning system. Therefore, the flow experience of computer-mediated communication postulated by Csikszenmihalyi (1998) might be more suitable to employ as a framework in online learning situations to examine elements of the AI TA that facilitate flow and presumably parasocial interaction in the future research.

The current work makes significant contributions to and has implications for education research and technological practices.

First, the study makes analytical contributions to the technology venturesomeness and anthropomorphism aspect in education research. Despite the hype surrounding Artificial Intelligence, the potential of it in relationship management remains underexplored in academia, let alone in the educational context. If anthropomorphism happens thoughtlessly, for the educational context, participants who were visible to a virtual human-identical teaching assistant should report a greater degree of perceived functional and social human-likeness toward the TA. Diffusion of Innovation is a useful perspective to understand the relationship between venturesomeness and adoption of AI TAs. However, the use of DI in the AITA studies was limited to recommending notional advices without empirical support. The present study is one of the first examples that empirically tested the DI being applied to understand the adoption of AI TAs. All the findings fully support the sustaining of the DI regarding the AI TA, more unambiguously in the AI-based tutoring and educational setting. That is, the DI can be useful to better comprehend how student adopt AI as TA in varied contexts, and it eventually develops the scope of the TAM.

Second, this current study examines attitudes related to pedagogical communication issues and ability to apply technology through the prism of personality-based psychographic aspect, thus operationalizing this theoretical basis in the light of venturesomeness and anthropomorphism. As stated earlier, while scientific capabilities are well progressive to create AI TA, there is little insights about how students would adopt the knowledge of AI TA. Recognizing that many educationalists remain doubtful about the worth of AI in the school, as well as the tendency that virtual agencies appear to support, the present study provides fundamental evidence that supports educators better comprehend how students may notice and respond to them.

Agreed that much research happens directing to the worth of AI in higher educational settings as it reports perceptive and communicative learning outcomes, possibly addressing undergraduate affect toward the adoption of AI TA in pedagogical context is the missing connection to persuade doubters and those that may currently feel apprehensive.

During pandemic and social distance, AI teaching assistants can support teachers by answering frequently requested questions from students. These queries, which occur throughout the semester and multiply in online classrooms with hundreds of students, may be a major undertaking for a teacher. Students benefit from the prompt delivery of AI TA replies.

REFERENCES

Bandura, A. 2001. Social Cognitive Theory of Mass Communication. Media Psychology, 3, 265-299.

Basinska, B. & Dåderman, A. 2018. Be fearless: Positive affect as a mediator between venturesomeness and self-efficacy in future entrepreneurs and managers. Revue Européenne de Psychologie Appliquée, 68, 171-180.

Belk, R. 2016. Understanding the robot: Comments on Goudey and Bonnin (2016). Recherche et Applications en Marketing (English Edition), 31, 83-90.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. & Tanaka, F. 2018. Social robots for education: A review. Science Robotics, 3, eaat5954.

Bloom, B. S. 1956. Taxonomy of educational objectives: The classification of educational goals. New York: David McKay Company.

Burgoon, J., Berger, C. & Waldron, V. 2000. Mindfulness and Interpersonal Communication. Journal of Social Issues, 56, 105-127.

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F. & Bull, S. 2013. Towards Empathic Virtual and Robotic Tutors. Artificial Intelligence in Education. International Conference on Artificial Intelligence in Education.
Choi, J. K. & Ji, Y. G. 2015. Investigating the Importance of Trust on Adopting an Autonomous Vehicle. International Journal of Human–Computer Interaction, 31, 692-702.
Chou, C.-Y., Huang, B.-H. & Lin, C.-J. 2011. Complementary machine intelligence and human intelligence in virtual teaching assistant for tutoring program tracing. Computers & Education, 57, 2303-2312.
Dautenhahn, K., Woods, S., Kaouri, C., Walters, M., Koay, K. & Werry, I. 2005. What is a robot companion - Friend, assistant or butler?
Doublerobotics 2017. Blended learning/hybrid classroom. Retrieved from https://www.doublerobotics.com/education/.
Edwards, A., Edwards, C., Spence, P. R., Harris, C. & Gambino, A. 2016. Robots in the classroom: Differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher". Computers in Human Behavior, 65, 627-634.
Epley, N., Waytz, A. & Cacioppo, J. 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. Psychological review, 114, 864-86.
Eysenck, S. B. G. & Eysenck, H. J. 1978. Impulsiveness and Venturesomeness: Their Position in a Dimensional System of Personality Description. Psychological Reports, 43, 1247-1255.
Gockley, R., Forlizzi, J. & Simmons, R. 2007. Natural person-following behavior for social robots. Proceedings of the ACM/IEEE international conference on Human-robot interaction. Arlington, Virginia, USA: Association for Computing Machinery.
Goetz, J. & Kiesler, S. 2002. Cooperation with a robotic assistant.
Goudey, A. & Bonnin, G. 2016. Must smart objects look human? Study of the impact of anthropomorphism on the acceptance of companion robots. Recherche et Applications en Marketing (English Edition), 31.
Guthrie, S. E. 1993. Faces in the clouds: A new theory of religion. Oxford Univ Press.
Guzman, A. L. & Lewis, S. C. 2019. Artificial intelligence and communication: A Human–Machine Communication research agenda. New Media & Society, 22, 70-86.
Hill, D. J. 2016. AI Teaching Assistant Helped Students Online—and No One Knew the Difference. SingularityHub, May 11.
Javaheri, A., Moghadamnejad, N., Keshavarz, H., Javaheri, E., Dobbins, C., Momeni, E., & Rawassizadeh, R. 2019. Public vs Media Opinion on Robots. arXiv Preprint, Retrieved from arXiv:1905.01615.
Johal, W., Castellano, G., Tanaka, F. & Okita, S. 2018. Robots for Learning. International Journal of Social Robotics, 10, 293-294.
Karl, M. 2018. Risk and Uncertainty in Travel Decision-Making: Tourist and Destination Perspective. Journal of Travel Research, 57, 129-146.
Kim, H.-J., Lee, J.-M. & Rha, J.-Y. 2017. Understanding the role of user resistance on mobile learning usage among university students. Computers & Education, 113, 108-118.
Kim, J., Merrill, K., Xu, K. & Sellnow, D. D. 2020. My Teacher Is a Machine: Understanding Students' Perceptions of AI Teaching Assistants in Online Education. International Journal of Human–Computer Interaction, 36, 1902-1911.
Kleijnen, M., Lee, N. & Wetzels, M. 2009. An exploration of consumer resistance to innovation and its antecedents. Journal of Economic Psychology, 30, 344-357.
Lee, Y.-H., Hsieh, Y.-C. & Hsu, C.-N. 2011. Adding Innovation Diffusion Theory to the Technology Acceptance Model: Supporting Employees' Intentions to use E-Learning Systems. Educational Technology & Society, 14, 124-137.
Legoherel, P., Hsu, C. & Daucé, B. 2015. Variety-seeking: Using the CHAID segmentation approach in analyzing the international traveler market. Tourism Management, 46, 359–366.

Lepp, A. & Gibson, H. 2008. Sensation seeking and tourism: Tourist role, perception of risk and destination choice. Tourism Management, 29, 740-750.
Li, J., Kizilcec, R., Bailenson, J. & Ju, W. 2016. Social robots and virtual agents as lecturers for video instruction. Computers in Human Behavior, 55, 1222-1230.
Masiero, L. & Qiu, R. 2018. Modeling reference experience in destination choice. Annals of Tourism Research, 72.
Mishra, L., Gupta, T. & Shree, A. 2020. Online teaching-learning in higher education during lockdown period of COVID-19 pandemic. International Journal of Educational Research Open, 1, 100012.
Moreale, E. & Watt, S. 2003. An Agent-Based Approach to Mailing List Knowledge Management.
Nass, C., Steuer, J. & Siminoff, E. 1994. Computer are social actors.
Newman, J. 2014. Apple's Siri Helps Autistic Teen With Communication Skills. npr.org, November 4.
Ozkan, S. & Koseler, R. 2009. Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation. Computers & Education, 53, 1285-1296.
Park, E., Kim, K. J. & Pobil, A. P. D. 2011. The effects of a robot instructor's positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom. Proceedings of the Third international conference on Social Robotics. Amsterdam, The Netherlands: Springer-Verlag.
Plog, S. C. 2002. The Power of Psychographics and the Concept of Venturesomeness. Journal of Travel Research, 40, 244-251.
Qiu, L. & Benbasat, I. 2014. Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. Journal of Management Information Systems, 25, 145-182.
Researchandmarkets 2017. Intelligent virtual assistant market by product, user interface, industry & geography - Global forecast to 2023. Retrievied from https://www.marketsandmarkets.com/Market-Reports/intelligent-virtual-assistant-market-184337393.html.
Rogers, E. M. 1995. The Diffusion of Innovations. Free Press: New York, NY.
Rogers, E. M. 2003. Diffusion of Innovations, 5th Edition, Free Press.
Rubin, A. M. & Perse, E. M. 1987. Audience Activity and Soap Opera Involvement: A Uses and Effects Investigation. Human Communication Research, 14, 246-268.
Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P. & Mohr, D. C. 2015. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res, 17, e175.
Sorokova, M. G. 2020. Skepticism and learning difficulties in a digital environment at the Bachelor's and Master's levels: are preconceptions valid? Heliyon, 6, e05335.
Tuomi, I. 2018. The Impact of Artificial Intelligence on Learning, Teaching, and Education: Policies for the Future.
Vanlehn, K. 2011. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46, 197-221.
Xu, D. & Wang, H. 2006. Intelligent agent supported personalization for virtual learning environments. Decision Support Systems, 42, 825-843.

DECISION SCIENCES INSTITUTE

Improving Efficiency in Wind Farm Maintenance and Service Processes

Sam Mangialardi
Illinois State University
Email: rsmangi@ilstu.edu

Bindu Gokila
Illinois State University
Email: bgokila@ilstu.edu

Kehinde Babatunde
Illinois State University
Email: khbabat@ilstu.edu

Borinara Park
Illinois State University
Email: bpark@ilstu.edu

ABSTRACT

Efficient servicing and maintenance of Wind Turbines (WT) is a top priority of wind farms throughout the United States. Currently, inefficacies in the delivery of components to service towers account for a significant portion of losses in turbine availability. This study uses the Discrete Event Simulation (DES) to simulate the current and suggested methods of component delivery and also suggests an appropriate addition to the service and maintenance operations to increase the efficiency of component delivery. The results of this study aim to assist wind farms of any size and location in the financial, operational, and efficiency problems associated with conventional component delivery processes.

KEYWORDS: Discrete Event Simulation, Renewable Energy, Wind Turbine Availability, Wind Turbine Downtime, Plant Efficiency, Service Cycle, Wind Turbine Component Failure

BACKGROUND

With an increase in global warming and limited fossil fuel reserves (Abramovskiy, Shalin, & Shestakov, 2019), wind energy has captured the commercial electricity-generating market in more than 80 countries. However, it is an expensive way to generate electricity (Bebi, Malka, Konomi, & Alcani, 2021). As per the 2016 Global Wind Operations and Maintenance Market Review Report, the global wind Operations and Maintenance (O&M) market is estimated to increase at a compound annual growth rate (CAGR) of 11.2% from $10 billion to $17 billion between 2015 and 2020. Research studies indicate a strong association between the cost of energy (COE) and O&M in this industry (Verma, 2012). Research data shows that O&M accounts for around 20-35 percent of this sector's total energy costs (Baagoe and Stentoft, 2016). Further research indicates that optimized O&M practices help reduce wind energy costs and increase safety (Nguyen et al.,2013).

INTRODUCTION

Global climate change and fossil fuel sustainability continue to be a driving force for renewable and non-renewable energy studies. While coal, oil and other non-renewables remain a focal point for large energy corporations, there remains an increasing need to explore renewable sources of energy like solar, wind and hydro to meet power consumption demand. This has resulted in a steady increase in wind turbine installation and the generation of wind energy globally. This increased investment in wind farms has created the demand for operations and maintenance(O&M) of various components installed at the wind farm (Chavan, 2019). As per the 2016 Global Wind Operations and Maintenance Market Review Report, O&M costs account for 10% of the newer wind turbines' annual cost and 35% for the older ones. Therefore, reducing O&M costs is a prioritized objective for wind farm stakeholders. Optimized O&M practices help create value and increase turbine availability by lowering downtime and increasing returns (Wallace, 2009). Wind turbines need to be producing electricity at peak performance rates whenever possible. When the wind turbine breakdown and stop producing electricity, the wind farm loses money. Minor component failures represent roughly 75% of all wind farm failures (Faulstich et al.,2011). The technician team's ability to repair these faults is critical, and the efficiency at which they make repairs is equally essential (Wallace, 2009). Lack of efficiency in delivering parts to affected wind turbines increases turbine downtime and decreases turbine availability. This results in production loss and increased maintenance efforts, resulting in unplanned downtimes up to 10 times per turbine per year (Faulstich et al.,2011). Nguyen et al. (2013), in their research work, uses semantic technology to facilitate data exchange through a data integration framework and improve O&M practice and reduce cost. Optimized O&M practices include fault analysis, optimized maintenance activities and visits, 24/7 alarm handling and analysis. Carlos et al. (2013) uses a stochastic model to select the optimum maintenance preventive strategy that minimizes the total cost and maximizes the annual energy produced. In this research study, maintenance frequency was considered as a decision variable. They used wind velocity probability distribution and Monte Carlo sampling methods in order to estimate the power generated.

Even though research studies are done to find an optimum strategy for O&M practice, the study made by Nguyen et al.,2013 using a semantic model is still a new concept in the wind industry. It needs more acceptance from wind industry stakeholders. In the research done by Carlos et al.,2013 each wind turbine is considered in isolation, and the O&M impact on the whole wind farm is not considered. Hence in both the models, the quantum of improving the efficiency of wind farm O&M process as a whole is not clear. The main purpose of this study is to focus on quantifying the efficiency of O&M process by addressing the problem of component delivery by using discrete-event simulation models.

CURRENT PROCESS

The Wind Farm simulation used for this research project is a representation of a Wind Farm located in a Mid-Western region in US. The Wind Farm includes 240 WT’s and employs roughly 15-20 Technicians during any given work week. During that week an average of 9 technicians will be scheduled to complete services. O&M has set a target goal of 12 services per 5-day week in an attempt to get back on target for their yearly service cycle. The process of delivering minor components to a faulted tower includes technicians receiving a work order, loading up the service tools and kit, driving to the tower, climbing the tower, and commencing the service. Once it is determined that the technicians will need an additional component(s), the process then includes climbing down the tower, driving to the O&M building, checking out and loading the component, driving back to the faulted tower, climbing the tower, and repairing/replacing the component. Occasionally, there will be someone available at O&M who can check out, load, and deliver the component to the faulted tower. This process is shown in our baseline simulation model, with a scaled number of turbines and technicians.

THEORETICAL DEVELOPMENT/ MODEL

A lack of efficiency in delivering minor components to affected WT's during scheduled services further increases downtime and decreases availability. This results in a lack of generation and significant loss of money and energy generated per Turbine. Despite the solutions provided in the reviewed literature, there is still a lack of reasonable solutions to the problem of component delivery efficiency. The existing bodies of the research work offer that utilizing an accurate stochastic model for individual turbine monitoring is an important factor in reducing turbine downtime and improving maintenance efficiency. However, they do not provide a reasonable suggestion for adequately removing the issue of component delivery inefficiency across the entire Wind Farm. Additionally, a semantic model to monitor and facilitate data exchange is suggested, but this too does not actually solve the problem of slow component delivery. Minor component failure results in a substantial reduction of wind turbine availability and increased turbine downtime. Utilizing the the Discrete Event Simulation features provided by the AnyLogic software, we are to specifically address the problem of component delivery efficiency. The key piece of data in these simulations is the lifespan of the simulation main agent, the service order. Because the length of time a service order is open is directly correlated with the amount of time a wind turbine is faulted during a service, we will be able to measure the total downtime of each serviced tower by studying the lifespan of the associated service order during both the previous delivery process and the improved process. Our simulation model proves that adding a mobile, centrally located parts delivery vehicle will help reduce the amount of time that technicians are waiting for replacement components in the WT as well as reduce the frequency of technicians leaving the tower and driving to O&M to retrieve the parts themselves.

PRE-INTERVENTION MODEL

Figure 1. Simulation model for taking baseline Service Order duration measurements

Shown in Figure 1 is a simulation model for the baseline service operation where a small wind farm of 10 WT's and two service teams of 2 service technicians. As a service work order is created and processed, it flows through the back-end process throughout the O&M building. This accurately represents the time and resources required to issue and carry out a service order. Once the order is surrendered to the service technician team, the team then takes the appropriate time to gather service supplies and travel to the wind turbine. Once at the WT, there is a probabilistic determination on whether or not an additional component will be required. If it is required, the technicians climb down the Turbine, leaving it running, drive back to O&M, consult with the Lead technician, locate and pack the required replacement component, load the component into the truck, drive back to the tower, crane up the part, and replace the part. Once the component has been replaced, the service team completes the service as normal and returns to O&M.

POST-INTERVENTION MODEL

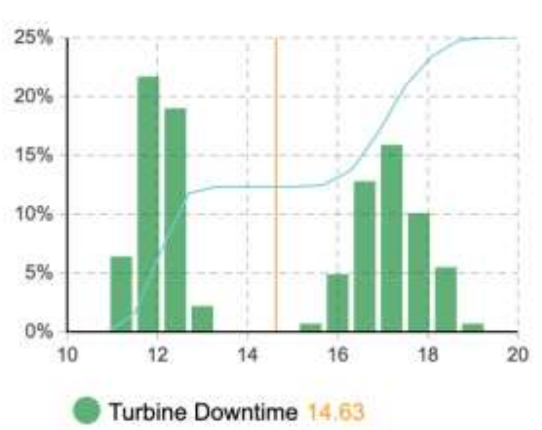
Figure 2. Simulation with the suggested improvements in the component replacement process.

Shown in Figure 2 is a simulation model to demonstrate the potential efficiency improvement by adding a mobile components delivery vehicle. Utilizing a centrally located and pre-loaded components vehicle eliminates the need for technicians to climb down the tower and drive back to O&M to locate and load their replacement component.

The proposed solution suggests creating a mobile parts and components vehicle that holds numerous amounts of the most commonly replaced minor turbine components. This vehicle would be centrally located inside of the wind farm and run parts to technicians as they are required and requested via radio or cell phone communication. This would not only speed up the time of repair and maintenance but would also eliminate confusion by keeping track of all components used through inventory recordkeeping and monitoring of components that may be refurbished. An initial experiment simulation determines the total downtime on a given eight-hour day with the previous methods of component delivery utilized. Gathering data to determine the percentage chance of a technician team needing a part brought to the turbine allows us to apply that percentage to a simulation and then simulate the technicians delivering the parts. Utilizing the simulation program allows us to identify the average downtime of each turbine using start and stop measurement allocations within the simulation itself. This allows us to measure the amount of time a turbine is offline due to the need for gathering replacement components. Additionally, measuring the lifespan of the service orders, which are directly tied to technician working hours, we are able to calculate the amount of time each turbine team spends servicing each turbine. With this information, we can calculate a total cost savings between each simulation model by applying current market statistics such as cost per mWh of electricity generated, wind farm size, and technician salary.

RESULTS

Both simulations ran for a total of 8,760 simulated hours each to represent one year of online generation of each wind turbine. Results were calculated using two different metrics. The first metric being the average amount of turbine downtime as a result of rendering the turbine offline during maintenance procedures. The Average Turbine downtime was measured with the use of the distribution function measuring the time from when the technicians first arrive at the turbine to right before they leave. This represents the technicians shutting down the turbine when they first arrive, and then starting the turbine up again just before they leave. Each output is tied to a histogram chart which gives us the visual data. The second metric used for data analysis was Technician Working Hours, calculated by the distribution of the life duration of each service order, which is directly tied with the working time of a team of two technicians. These two metrics are shown in Figure 3 and Figure 4, respectively, for the pre-intervention model. And Figure 5 and Figure 6 show the metrics for the post-intervention model. Figures 3 and 5 show a mean turbine downtime of 14.63 hours and 13.23 hours respectively. Figures 4 and 6 show a mean of 777.17 and 568.87 hours of Technician Hours Per Order respectively.



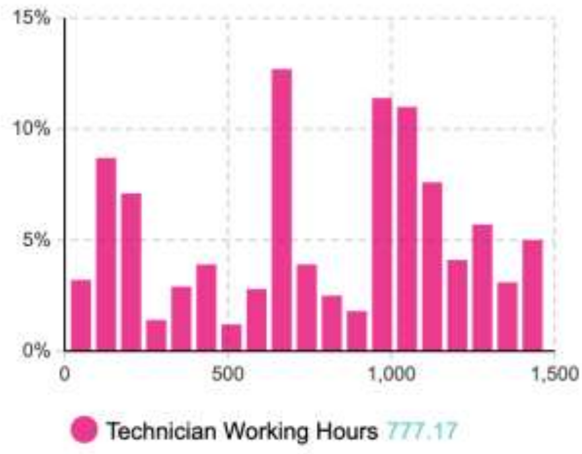
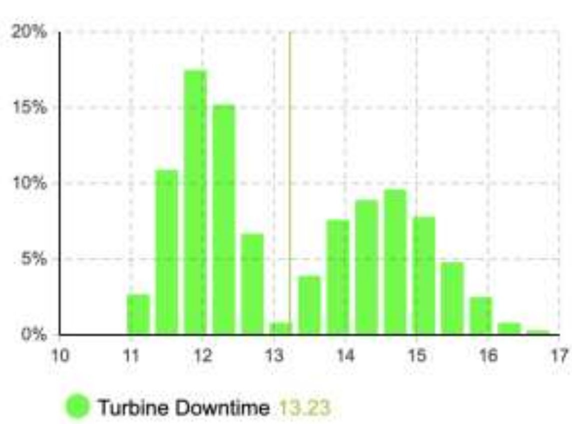


Figure 3. Average Turbine Downtime in Hours Pre-Intervention (left) & Figure 4. Average Technician Working Time in Hours Pre-Intervention (right)



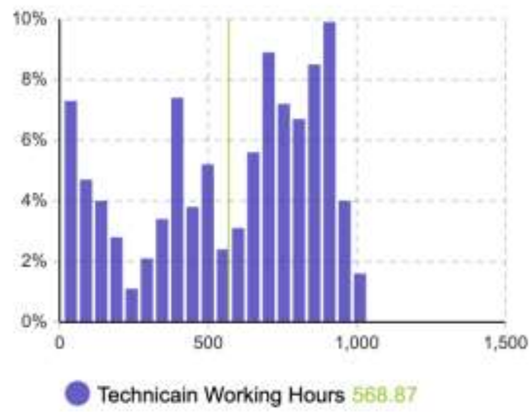


Figure 5. Average Turbine Downtime in Hours Post-Intervention (left) & Figure 6. Average Technician Working Time in Hours Post-Intervention (right)

CONCLUSION

The proposed Discrete Event Simulation solution addresses the inefficiency in current component delivery methods to faulted turbines during regularly scheduled service and maintenance procedures. While some research exists which acknowledges this problem, there is currently no practical solution to improve the significant amount of downtime that results from technicians being required to stop work, climb down the wind turbine, and drive back to O and M to locate any replacement parts needed during their workday. Similarly, wasted man hours has yet to be addressed as a significant source of lost capital. In this study, the proposed solution was able to decrease turbine downtime by an average of 1.4 hours or 9.57% total downtime hours per turbine as well as reduce technician labor hours by 228 hours per turbine per year. On a medium-sized, 30,000 mWh annual output wind farm utilizing 240 wind turbines at maximum capacity, the addition of a mobile components vehicle resulted in a substantial cost savings of $10,080,000 per year by reducing turbine downtime alone.

With a service team of two technicians, the solution resulted in a reduction of labor hours resulting in a cost savings of $2,497,000 per year on a 30,000 mWh wind farm with 240 turbines. The combined total of $12,577,000 saved per year by increasing operations efficiency is a substantial amount of potentially lost revenue for a wind farm. The calculations were based off the current Levelized Cost of Energy for wind generated electricity (Girouard, n.d), the average hourly salary of wind turbine technicians in the U.S (Severson, 2017), and the energy output of the Vestas V164 Wind Turbine (Varrone, 2018), which generates 30,000 mWh per year.

Implementing the use of a mobile components delivery vehicle has the potential to save wind farm operators nearly enough money to construct one - 7 mWh turbine every year with the capital saved alone. The size of the wind farm used in our calculations was based on local wind farm sizes around a Mid-Western region in US. Wind farms across the United States range in size and amount of employed technicians and these calculations can easily be adjusted to provide cost saving analyses of any wind farms with larger turbine capacity and employee size.

LIMITATIONS & FUTURE RESEARCH

Staffing of technicians varies throughout the workday and work week. It is required that at least two technicians be at a WT in order for it to be serviced. It is common for there to be fewer technicians than is needed to complete a service. Employee illness, injury, or other repair priorities are among the reasons for a limited amount of service staff on any particular day. These factors are harder to predict and not included in the simulation

Individual work responsibilities vary and there are certain circumstances that prevent technicians from being able to service turbines. These circumstances include weather related work closures, department training days, in-house maintenance days. If these occur unexpectedly, it disrupts the normal servicing schedule and creates a lag in normal service cycling. These are not included in the simulation.

With more accurate and parameter variations as well as stronger real world data collection techniques it is likely we would see the gap in turbine downtime length continue to increase. Similarly, a stronger representation of a real world 2-year service cycle would return even more accurate results.

Potential delivery options include multiple mobile component delivery vehicles and drone delivery. Utilizing drone technology to deliver components within the capabilities of current market drones may significantly increase the efficiency of component repair and replacement during regularly scheduled maintenances.

REFERENCES

Abramovskiy, A., Shalin, V., & Shestakov, S. (2019). Wind energy development policy as a type of alternative energy source. *Journal of Environmental Management & Tourism*. doi:doi:http://dx.doi.org.libproxy.lib.ilstu.edu/10.14505/jemt.v10.5(37).01

Baagoe, V., & Stentoft, J. (2016). Operations and maintenance issues in the offshore wind energy sector. *International Journal of Energy Sector Management*, 245-265. doi:DOI:10.1108/IJESM-04-2015-0012

Bebi, E., Malka, L., Konomi, I., & Alcani, M. (2021). An analysis towards a sustainable energy system in albania highly supported by large scale integration of wind resources: (A case study of mamaj wind farm). *International Journal of Energy Economics and Policy*, 355-372. doi:Retrieved from http://libproxy.lib.ilstu.edu/login?url=https://www-proquest-com.libproxy.lib.ilstu.edu/scholarly-journals/analysis-towards-sustainable-energy-system/docview/2501473860/se-2?accountid=1157

Carlos, S., . Sánchez, A., Martorell, S., & Marton, I. (2013). Onshore wind farms maintenance optimization using a stochastic model. *Mathematical and Computer Modelling, 57*(7-8), 1884-1890. doi:https://doi.org/10.1016/j.mcm.2011.12.025.

Faulstich, S., Hahn, B., & Tavner, P. (2011). Wind turbine downtime and its importance for offshore deployment. *Wind Energy, 14*. doi:10.1002/we.421

Girouard, C. (n.d.). The numbers are in and renewables are winning on price alone. Retrieved May 12, 2021, from https://blog.aee.net/the-numbers-are-in-and-renewables-are-winning-on-price-alone

GlobalData, UK. (2016). *Global Wind Operations and Maintenance Market Review.* Retrieved from https://store.globaldata.com/report/gdae0257vpt--global-wind-operations-and-maintenance-market-review/

Nguyen, T. H., P. A., Friiso, T., Nossum, R., & Tyapin, I. (2013). A framework for data integration of offshore wind farms. *Renewable Energy, 60*, 150-16. doi:http://dx.doi.org.libproxy.lib.ilstu.edu/10.1016/j.renene.2013.05.002

Severson, D. (2017, November 21). How much money does a wind turbine technician make per year? Retrieved May 12, 2021, from https://work.chron.com/much-money-wind-turbine-technician-make-per-year-18571.html

Varrone, C. (2018, February 21). Meet a Gargantuan wind turbine, THE 7-Megawatt V164. Retrieved May 12, 2021, from https://www.theatlantic.com/technology/archive/2011/05/meet-a-gargantuan-wind-turbine-the-7-megawatt-v164/238662/

Verma, A. (2012). Performance monitoring of wind turbines: A data-mining approach. doi: Retrieved from http://libproxy.lib.ilstu.edu/login?url=https://www-proquest-com.

Wallace, J. (2009, January). *Wind turbine downtime and the O&M team*. Retrieved from http://www.renewableenergyfocus.com/: renewableenergyfocus.com/view/1397/wind-turbine-downtime-and-the-o-m-team/

DECISION SCIENCES INSTITUTE
Analyzing Sustainability of Cities using Design Science and Location Analytics

Neelam Raigangar
Claremont Graduate University / University of Redlands
Email: neelam_raigangar@redlands.edu

James Pick
University of Redlands
Email: james_pick@redlands.edu

Avijit Sarkar
University of Redlands
Email: avijit_sarkar@redlands.edu

ABSTRACT

Unplanned urbanization leads to social, environmental, and economic deterioration. To mitigate the impact, cities must be sustainable and use information and communications technologies to build institutional capacity. The objective of this study is to utilize design science to establish a framework for a sustainable city that includes methodology, artifacts, sustainability indicators, and theories. The research output can be used as a repeatable process by cities to become sustainable. To visualize and analyze changes, Location Analytics is leveraged to gauge how successful City of Los Angeles, as a sustainable city, is in meeting UNs Agenda 2030 sustainable development goals.

KEYWORDS: Sustainability, GIS, Location Intelligence, Location Analytics, Design Science Research

INTRODUCTION

Estimates are that by 2050, 66% of the global population will be urban (UN, 2014). This continuous urbanization process will pose a challenge to urban life if not planned well. Challenges include, but are not limited to, higher rate of energy consumption, decline in public health, increases in traffic congestion, worsening air and water pollution, increases in toxic waste disposal, resource depletion, and social inequality and vulnerability. Urban and academic circles have been debating on how to respond to these challenges and the unsustainability of the current urban environment. (Bibri et al., 2017).

The United Nations (UN) suggests that governments must implement policies to focus on sharing equitably and sustainably the benefits of urban growth and recommends a need for a holistic planning approach as the cities are the conduit that can “lead the way towards economically, socially, and environmentally sustainable societies.” (UN, 2014, p. 17). Sustainability is important to mitigate social, environmental, and economic downgrade impacts of unplanned urbanization.

For a sustainable urban environment, there is need to apply integrated approaches to build the institutional capacity by using information and communications technology (ICT). With continued urbanization in recent years, cities are using ICT’s to become sustainable cities. The investment and research in ICT's, motivated by the desire for economic benefits and social gains, will allow cities to use these technologies to achieve sustainability, reliability, energy efficiency, and productivity.

Sustainable cities have grown in importance and are vying to be regional standouts to make them lucrative for people to reside there and attract investments in the development of the city. Eric Garcetti, the mayor of LA declared: "In Los Angeles, sustainability is a core value that guides all of our work, because our survival depends on it." (Mayor's Office Los Angeles, 2019). The update to the 2015 Sustainable City Plan identified socioeconomic and environmental elements where LA has been successfully in leading the nation in renewable energy, reducing water use, moving to zero emissions, increasing electric vehicle chargers, etc. (Garcetti, 2021). Hence for the purpose of our research we have focused on the City of Los Angeles (LA).

The City of LA "is committed to using IoT to boost environmental awareness, livability, and mobility. (Stone, 2018)" Government Technology Magazine has twice named Los Angeles as the #1 Digital City in the US at the annual League of Cities conference for city mayors and council members. "The city's initiatives apply IoT to improving its citizens' safety and lifestyle, the environment, and saving people time." (Stone, 2018). City of Los Angeles along with University of Southern California and other companies, government entities, non-profits, and academic institutions have formed the Intelligent IOT Integrator (I3) consortium that is focused on "…creating a vibrant and rich IoT community centric data ecosystem". (USC, 2021). The other reason we focused on the City of LA is that: "…since 2013, Los Angeles has progressed from a moderate digital status to a leading digital city." (Pick, 2016).

Research conducted by Woetzel et al., (2018 pg.12) for McKinsey concluded that the City of LA ranks 2nd in the U.S. and has scored 34.5 points out of maximum of 55 for deployment of sustainable city applications involving mobility, security, utilities, healthcare, economic development, housing, and community. As the City of LA is heavily invested towards sustainable city applications, we see improvements over time in some sectors and in other sectors there is definitely potential for improvement. As per Stone, (2018), "One of those standout cities is Los Angeles. The second largest American metropolis, the City of Los Angeles is committed to using IoT to boost environmental awareness, livability, and mobility." The City of Los Angeles is also a leader in applying location analytics to explore and solve city problems. Location analytics is applying data mining, visualization, machine learning, and other analytics methods to geographically-referenced data in order to better understand spatial processes.

The remaining sections of this study are structured as follows. The next section states the research purpose and research questions and justifies the findings. In the Contribution section, we indicate our research contribution to MIS literature and expound how our suggested frameworks and DSR methodology and artifacts can be leveraged by other cities to measure their success level towards becoming sustainable cities. In the literature review section, we focus on analytical and visualizing tools and sustainability. A subsection describes how, for cities, we shortlisted six indicators from UN SDG's since they were a match to Genuine Progress Indicators (GPI) that US states use to analyze their state's progress. These indicators were also identified by US city mayors as relevant for their city's progress.

We then explain elements of the Technology-Organization-Environment (T-O-E) theory that informs our study for technology adoption. The Research Methods section describes our framework design, DSR methodology, and artifacts that are the outcome of this study. We analyze a multi-attribute locational analytics artifact which is useful to visually analyze areas in the City of LA. We map the social, economic, and environmental sustainability indicators that have changed over time and display the percentage of the change to highlight areas that have been successful or need attention.

In the Research Findings section, we apply location analytics to examine sustainability changes 2014-2018 for the City of Los Angeles. Location analytics is applied to summarize results of the indicators that have been mapped, and observations are provided of the sustainability of the City of Los Angeles. In the same section, for Design Science analysis, we focus on artifact development and refinement, and the analysis garnered utilizing the refined artifacts.

For the Discussion, implications, limitation and future research section, we connect our findings to the study's objective and research questions and discuss research implications and study limitations and future research possibilities. Finally, the Conclusion section shows the importance of our approach to measure sustainability success using DSR and location analytics and suggests following our frameworks and repeatable methodology to instantiate and design artifacts to analyze other smart cities and their sustainability indictors.

RESEARCH PURPOSE AND CONTRIBUTION

Objective and Research Questions

The objective of this study is to utilize design science methodology (Hevner et al., 2004; Wieringa, 2014) to generate new artifacts or notably improve existing artifacts with regard to use of ICT towards the goal of improving sustainability

Our study also focuses on two research questions, the findings of which will provide valuable guidance to other cities in terms of providing them with a method that is repeatable and is iterative in nature as they design artifacts relevant to their sustainability goals and targets.

> RQ 1) How essential is location analytics to visualizing and analyzing sustainable cities success in meeting the socioeconomic and environmental sustainability goals/targets of cities?

The first research question relates to technology for analyzing and visualizing the data. Analyzing data over time is essential for decision making. Using appropriate technology to understand the visual depiction and making visual sense of the data is also critical to decision making. Since our study is focused on indicators at country, state, and city levels, which are all geographic entities, we have used location analytics as the analytical and mapping tool for our research to assess change and uncover opportunities for improvements. "… location data provides unique insights, revealing hidden relationships, patterns, and trends that drive stronger decision-making." (Esri, 2021). We map areas that have seen a change over time and visualize the percentage change by zip code for sustainable development goals (SDG). Location analytics is an artifact that can be leveraged for decision making for long term investments in the development of the city.

> RQ 2) What are the artifacts produced by a design science framework for attaining city socioeconomic and environmental sustainability goals?

The second research question relates to the application of design science framework to produce artifacts that have been iteratively developed and refined. By using the artifacts for the instantiation methodology, frameworks, and visualization tools, cities can gain insights from problem areas and areas that are showing improvement. Decision making can be enhanced by studying patterns and trends that are a guidepost to attaining the desired goals. The elements from the theory can be adopted and used as a benchmark to achieve a successful outcome. These artifacts can be a guide to cities desiring to be successful to attain their socioeconomic and environmental sustainability goals.

Contribution

Our research contribution to the MIS knowledge base is to posit design frameworks and design methodology that can be applied by other cities desiring to be successful in meeting sustainability goals/targets and identify areas for improvement.

We propose a research framework that can be used as a repeatable process to inform future research that can take advantage of our findings, resulting in comparable studies for other cities. The research framework helps to identify sustainability targets and apply elements from theory for technology adoption. The framework constitutes a DSR methodology since it is a problem-solving paradigm and supports refinement of artifacts based on iterative development.

Since our research is dependent on ICT, technology adoption is a key to success in implementation and instantiation of a sustainable city. Furthermore, basing the study on elements of T-O-E theory is a contribution to the MIS knowledge base since other cities can leverage the elements of the theory identified for technology adoption to achieve socioeconomic and environmental sustainability in a city.

Our practical research contribution is to adjust and apply the DSR methodology to improved sustainable design for the City of LA. DSR is a problem-solving paradigm, is informed by theory, and allows for iterative process for artifact development and refinement. Hevner et al. (2004) identify software prototypes as an artifact which helps to understand and solve a problem. By applying DSR methodology to the City of LA, our contribution can guide other cities in instantiating a sustainable city and to benchmark and be on par with existing sustainable cities.

Finally, the artifact for visualizing and analyzing data is location analytics. GIS allows city government to manage, visualize, and analyze data to gain insights by asking questions about the data’s surroundings and studying maps that summarize complicated data and themes. With location analytics cities can visualize changes over time and identify spatial patterns and relationships between key variables for a given location and visualize and analyze layers for social, economic, and environmental indicators for decision support.

LITERATURE REVIEW

With rapid urbanization, cities are facing challenges such as higher rate of energy consumption, decrease in public health, increase in traffic congestion, worsening air and water pollution, toxic waste disposal, resource depletion, and social inequality and vulnerability. To support this continuous changing flow of people, cities have to be sustainable. UN SDG Agenda 2030 have identified 17 goals with a total of 168 targets for sustainability.

Sustainability is important to mitigate social, environmental, and economic deterioration impacts of unplanned urbanization. Investment and research in information & communication technologies (ICT) motivated by a desire for economic benefits and social gains will allow cities to use these technologies to achieve sustainability, reliability, energy efficiency, and productivity by applying integrated approaches to build institutional capacity and achieve their sustainability goals.

Sustainable City

Sustainability is defined as being able to continue to grow presently without impacting the needs of the future generations to come (WCDE, 1987). Hoornweg (2016) defines sustainable development as "…the nexus of wealth generation, economy, equity, environmental degradation, urbanization, well-being, culture, creativity, and local and global governance. To a large extent, sustainable development is driven by cities, the way they are built and managed, and the way people live in them." (p. 3). A sustainable environment ensures efficient resource usage thus ensuring resources last longer.

<u>Identifying Sustainable indicators for Social, Economic, and Environmental Domains</u>

Since a city is a nexus of social, economic, and environmental domains, it becomes important to ensure sustainability in these domains and understand the sustainability indicators for them so that they could be assessed, monitored, and can inform the concentrated effort on planning for continued sustainable growth. To identify the indicators, we started off broadly at the global/country level by looking at UN SDG target for countries, then narrowed it down to indicators that are relevant at state level by studying the Genuine Progress Indicators (GPI) that states use to measure progress, and further narrowed it down to indicators identified by city mayors as important to measure success of their cities (Table 1).

<u>Global/Country Level Indicators</u>

To identify indicators relevant to our research, we reviewed UN Sustainable Development Goal's (SDG) Agenda 2030 and their targets. UN aspires that these will be adopted by all countries to meet sustainability goals by year 2030 (UN, 2015). UN SDG has identified 17 Goals and 168 targets of which this study identified six targets from social, economic, and environmental domains that are relevant at state and city levels as well.

<u>State Level Indicators</u>

Multiple states in the US use Genuine Progress Indicators (GPI) instead of Gross Domestic Product (GDP) to understand their state's progress. Daly et al. (1989) came up with Index of Sustainable Welfare (ISEW) as a welfare measure which was later named Genuine Progress Indicator. GPI looks at social, technical, and economic development to measure progress and looks at the externalities – benefits and costs of production across social, economic, and environmental domains. On the other hand, GDP looks at production and not costs i.e. excluding negative externalities such as environmental cost of pollution, depreciation, costs of standard of living, etc. States such as Hawaii, Utah, Colorado, Maryland, Ohio, Vermont have used GPI to measure progress (Bagstad et al., 2014). GPI is used to influence policy development for improved sustainability and to ensure continued progress of their state (Bagstad et al., 2014).

City Level Indicators

A survey of city mayors conducted by United States Conference of Mayors (2018), indicated they were interested in similar indicators for social, environmental, and economic development to ensure success of their cities.

In Table 1, we have identified UN sustainable development targets that are related to our study at the country, state, and city levels. We provide a comparison of UN targets, GPI, and city-level indicators from mayor’s survey.

Table 1: Indicators for social, economic, and environmental domains

UN Sustainable Development Targets (Country Level) which are related to this study	**GPI (State Level)**	**Mayors Survey (City Level)**
• Promote development-oriented policies that support productive activities, decent job creation, entrepreneurship, creativity and innovation, and encourage the formalization and growth of micro, small and medium-sized enterprises, including through access to financial services. (UN Goal 8.3)	Social Indicator – Underemployment /Higher Education	Creating Jobs
• Develop quality, reliable, sustainable, and resilient infrastructure, including regional and transborder infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all. (UN Goal 9.1)	Economic Indicator – Income Inequality/Poverty/ Net Capital Investment/Personal Consumption	Attracting Private Investment
• By 2030, reduce the adverse per capita environmental impact of cities, including by paying special attention to air quality and municipal and other waste-management. (UN Goal 11.6) • Integrate climate change measures into national policies, strategies, and planning. (UN Goal 13.2)	Environmental Indicator - Air/Water Pollution/Pollution Abatement/Climate Change	Improving Energy Efficiency

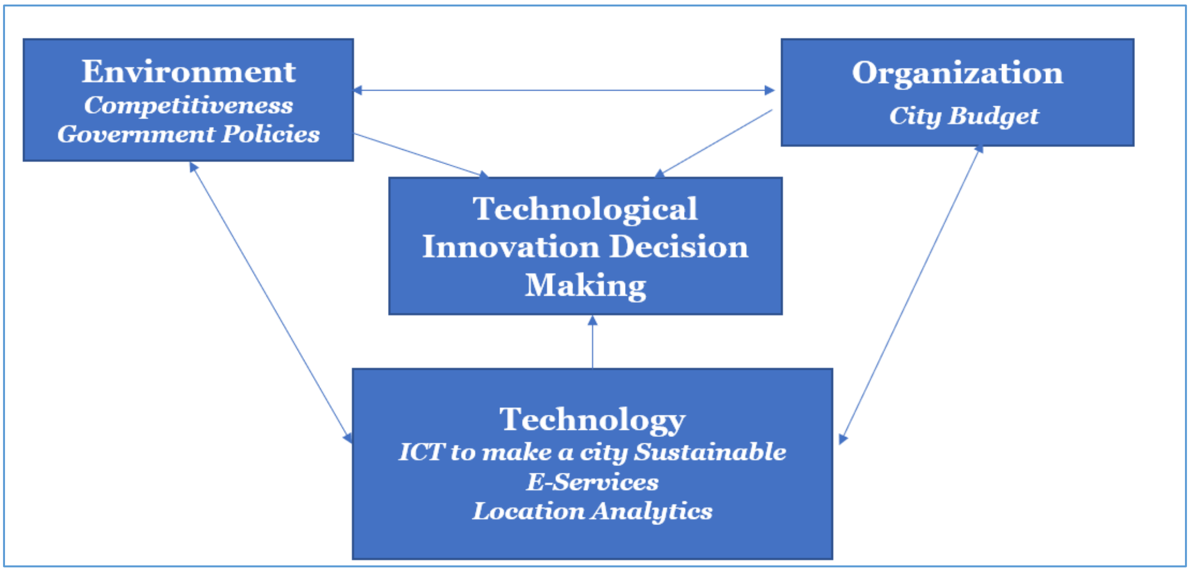
Theoretical Background: Technology-Organization-Environment (T-O-E) framework

The T-O-E framework (Tomatzky and Fleischer, 1990; Baker, 2011) was developed to study the adoption of technological innovation. Lippert et al (2006), identified multiple empirical studies that used T-O-E for analyzing the acceptance by organizations of new technologies (Chau and Tam, 1997; Gibbs and Kraemer, 2004; Lacovou et al., 1995; Kuan and Chau, 2001; Thong, 1999; Zhu et al., 2004; Zhu et al, 2003; Zhu and Kraemer, 2005).

In the framework, technological, organizational, and environmental contexts influence how an organization adopts and accepts new technologies (Lippert et al. 2006). Technological context focuses on internal and external technologies that will aid in organizational productivity. Organizational context focuses on the resources for innovation that are available, such as firm size and scope; the centralization, formalization, interconnectedness, and complexity of its managerial structure; and the quality and availability of the firm's human resources.

Environmental context concentrates on the external setting that business is in and what forces influence the industry, the competitors, resources, and governmental interactions. Since ICT is key to a sustainable city, technology adoption is critical. There are elements from T-O-E framework that are applicable for technology adoption by cities looking to match or better the City of Los Angeles in becoming sustainable cities (Figure 1).

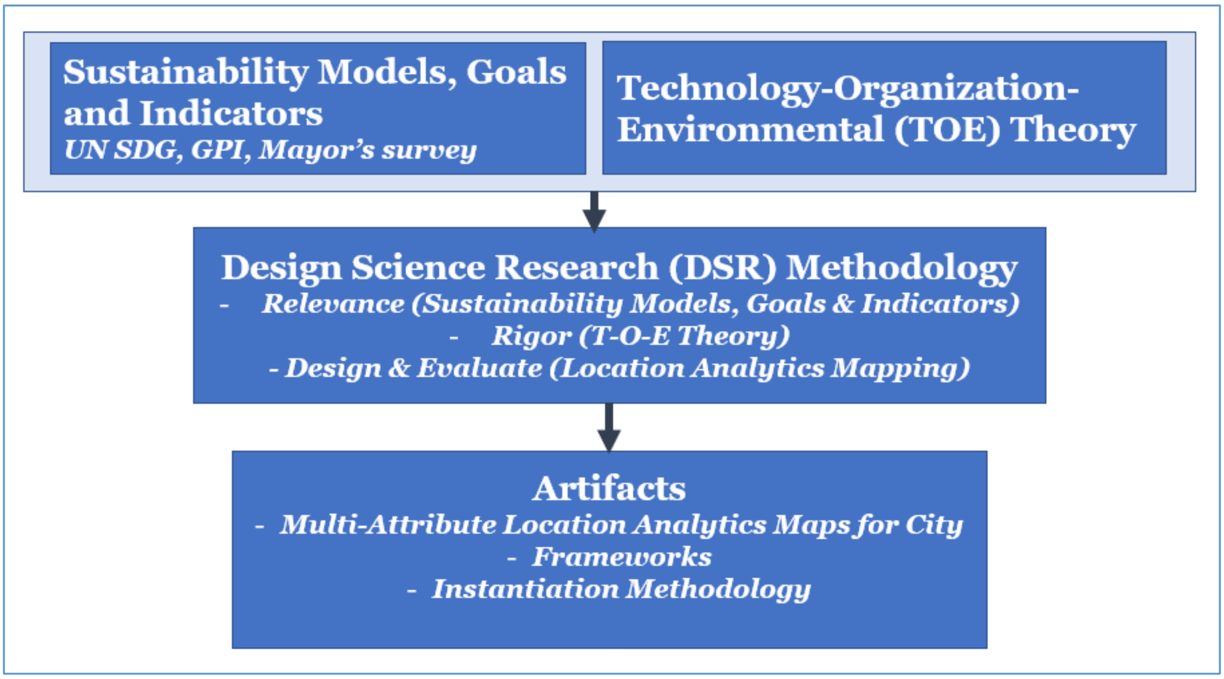
Figure 1: T-O-E elements for the City of Los Angeles



RESEARCH METHODS

We follow a research framework which is summarized in Figure 2. Our research framework details steps that helped us conduct this study. We explain our framework below.

Figure 2: Research framework



Technology-Organization-Environment Theory (T-O-E)

A sustainable city is dependent on leveraging ICT's. Hence for the city to become sustainable, there is a need for technology adoption. T-O-E elements identified in our research can help the city with technology adoption that is imperative. T-O-E informs the DSR methodology as well. With T-O-E, cities can use elements of the theory for their own technology adoption to move towards a sustainable city.

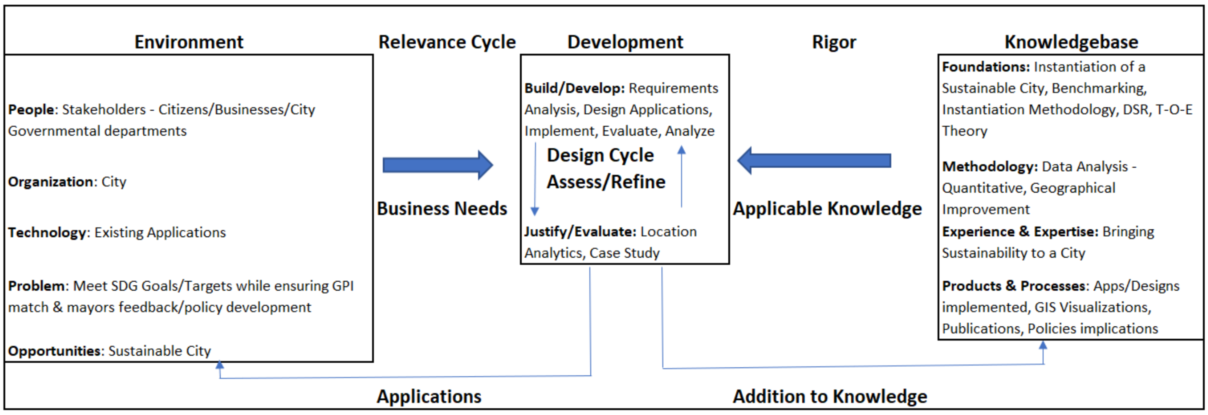
Applying the TOE framework to City of LA - the technology investments are influenced by the interplay of environment, organization, and technology elements. Leveraging each of the elements will allow to make LA a sustainable city and provide the blueprint for other cities to emulate and benchmark. The environment is the city's policies that they formulate to attract educated people, beat competition, as well as attract investments to grow a city into a profitable city. By formulating policies leading to job growth and encourage esteemed universities to set up campuses, the city can leverage this element to their benefit. For the organization element, the city can budget and provide resources to support technology investments. Identifying technology gaps and investing in technology that would provide e-services would make it convenient to do business with the city. Investing in location analytics and visualization tools and other ICT's can aid with identifying areas that are sustainable or need attention.

Design Science Methodology

DSR is a problem-solving paradigm. It is informed by theory. DSR supports an iterative process for developing and delivering artifacts. Since we are trying to assist cities deal with challenges of meeting UN SDG's, we adopt DSR as our research methodology as it will allow them to benchmark and use expertise and experience to develop into a sustainable city. As per Chatterjee (2015) “The design process is a sequence of expert activities that leads to the creation of an innovative product (artifact). The evaluation of the artifact provides feedback and better understanding of the problem to improve quality of the product and the design process.”

The 3 iterative cycles for DSR (Hevner et al., 2004) shown in Figure 3, are: 1) *Relevance Cycle*: There is a problem and an opportunity to find a solution for the problem. In the present study, the organizational environment plays a key role since the research is impacted by people, organizations, and technology that necessitate the need for the research project. 2) *Rigor Cycle*: To design the process/product, we examine the existing knowledgebase such as theories, frameworks, instruments, constructs, models, methods, and instantiations as well as the methodologies such as data analysis, formalisms, measures, and validation criteria. We apply experience and expertise to design rigorous processes and products. 3) *Design and Evaluate cycle*: This cycle focuses on building theories/artifacts. It uses analytics, case studies, experiments, field studies and/or simulations for evaluation. Evaluation allows the investigator to refine theories and artifacts before final release. Finally, the output of the research adds to the knowledge base and applies the design for the benefit of the organization.

Figure 3: DSR framework



The specific design science steps are to address the evaluation of whether or not, the City of LA meets UN SDG targets specific to indicators relevant at state and city level. The steps consist of the following:
1) Relevance Cycle for the City of LA. Examine our environment that is the city and its residents, businesses, and governmental entities. The problem is to meet the UN targets by 2030. We have the opportunity to make the City of LA sustainable by using ICT.
2) Rigor Cycle for the City of LA. Examine the knowledgebase level and determine how to improve upon existing processes/policies for the betterment of the City of LA residents. We can leverage people who have prior experience and are experts in this field to assist with setting up a sustainable city. This includes examining elements of T-O-E theory, design frameworks, and the DSR approach.
3) Design and Evaluate Cycle for the City of LA. The steps of the systems development lifecycle phases are assessed and refined. We worked through an iterative process for the multi-dimensional location analytics maps to have a visualization that succeeded in displaying indicators for the City of LA appropriate for our study. We worked with various color schemes

and icons, data sets and map layers for different time periods to assess change over time and pinpoint areas for improvement.

To maximize value from location analytics, we follow the 4 imperatives for it, as recommended by Esri (Esri, 2013):
Imperative #1-Go Beyond Basic Mapping. Rather than just mapping dots, use tools for automated clustering, heat mapping, data aggregation, color coding, etc. For our study, we used the percentage change and color coding to see areas impacted over time.
Imperative #2—Enrich Your View. Using additional geo-enriched data layers such as demographic data, tapestry segmentation, third party data, etc. The geo-enriched data we used were sourced from California Office of Environmental Health Hazard Assessment (2014, 2018). Demographic and socioeconomic attributes from the US Census American Community Survey data (US Census, 2021) were used for enrichment as well.
Imperative #3—Perform Map-Driven Analysis. Mapping the data can uncover many patterns and insights. Applying spatial queries and spatial modelling can provide insights about relationships between socioeconomic variables and environmental impacts within a city. We conducted location analysis for the City of LA and mapped areas that have Los Angeles, CA zip codes as geographic units and does not include San Fernando Valley or areas extending down to Long Beach. This analysis led us to identify areas that were impacted over time both positively and negatively for our indicators.
Imperative #4—Collaborate with Maps. Sharing these maps allows collaboration between various stakeholders and raises awareness. The detailed layers visualized on a map help with the identification of local conditions that support sustainability. We further built dashboards to visualize patterns and understand changes in our study area and associated data over space and time.

Artifacts

Multi-Attribute Location Analytics Maps

Location intelligence is powered by geographic information system (GIS) and is rooted in science. GIS allows to manage, visualize, and analyze data to gain insights by asking questions about the data's surroundings and studying informative maps that summarize complicated themes, spatial statistics, and analytics. "… location data provides unique insights, revealing hidden relationships, patterns, and trends that drive stronger decision-making." (Esri, 2021). Location analytics helps to harness the power of geospatial data to uncover spatial patterns and trends, examine spatial relationships between key variables, and provides insights for decisions and actions. The iterative process has been instrumental in successfully delivering and displaying appropriate color schemes, icons and layers for the artifact. Maps of the City of LA display sustainability indicators for 2014 and 2018 and indicate percentage change from 2014 to 2018.

Frameworks

As one of the artifacts, we have designed and identified elements for the following frameworks:
'Figure 1 – T-O-E framework' – This framework describes T-O-E elements that can help the city with technology adoption taking into consideration the environment, the organization and technology. This will aid with decision making on technology adoption to support sustainable city growth/development efforts.

'Figure 2 - Research framework' – This framework guides a city that is desirous of being a sustainable city. The city can utilize our suggested framework to become sustainable and enabling it to meet UN SDG Agenda 2030 targets.

Instantiation Methodology

We have provided 'DSR framework' (see Figure 3) that can be used to set up an instance of a sustainable city. DSR as a problem-solving paradigm can be useful for a city desiring to follow an established methodology. The methodology allows for opportunities to resolve a problem, understand the stakeholders impacted, and technology that can be implemented as a solution to mitigate the problem. Existing knowledgebase is applied, and data analysis is conducted for evaluation purposes. It encourages artifact refinement due to the iterative cycles that are part of the design cycle. In a nutshell, theory informs the methodology so that the elements can be applied for technology adoption and benchmarking for the city to stimulate its development.

RESEARCH FINDINGS

Location Analytics – Spatial Sustainability Changes 2014-2018 in City of Los Angeles

Mapping provides valuable spatial intelligence. At a glance we can see differences over time as we map layers for the indicators. In the dashboards that we build for location analytics, we are able to click on zip codes on the map and see the metrics for them. In this paper, we are limited to displaying mapping visualization to track spatial variation over time. The visuals allow us to analyze the impact of the metrics for those indicators. City planners and policy makers can avail of the spatial intelligence to plan and target areas that are lagging or have potential for further improvements.

We have provided maps for sustainability indicators – unemployment, education, poverty, pollution, toxic release, and solid waste maps (Figures 4a to 4h). For unemployment indicator (Figures 4a-4c), we have given slightly enlarged maps to show 2014 and 2018 numbers for unemployment in the City of LA and the percentage change from 2014-2018. Maps for the remaining indicators, for the same time period, have been miniaturized because of DSI's file size limitations. Analysis of each indicator has been provided under each map visual.

Figure 4: Location analytics-based dashboard with Sustainability Indictors, City of LA (2014-2018)

Figure 4a: City of LA – Indicator: Unemployment. (2014)

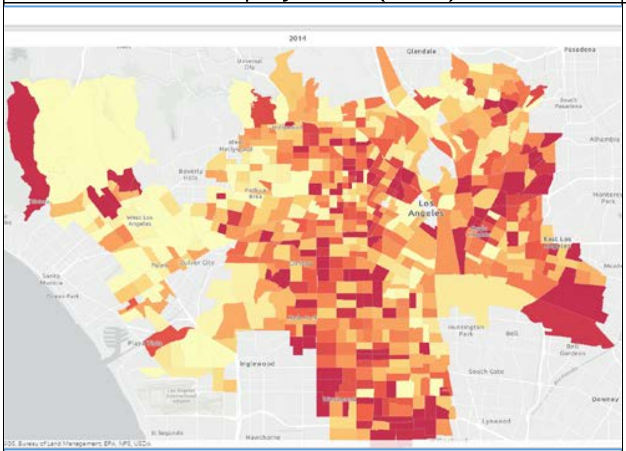


Figure 4b: City of LA – Indicator: Unemployment. (2018)

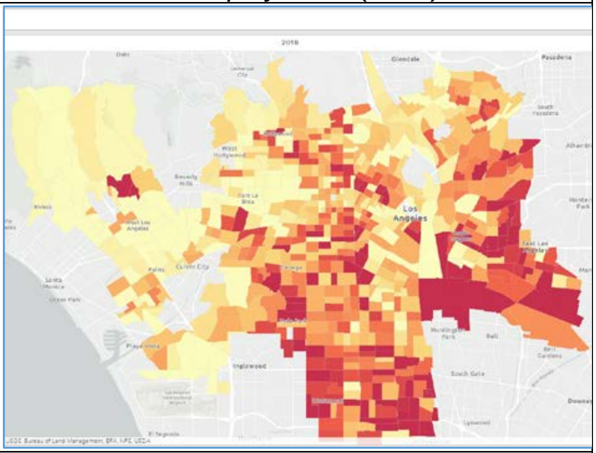
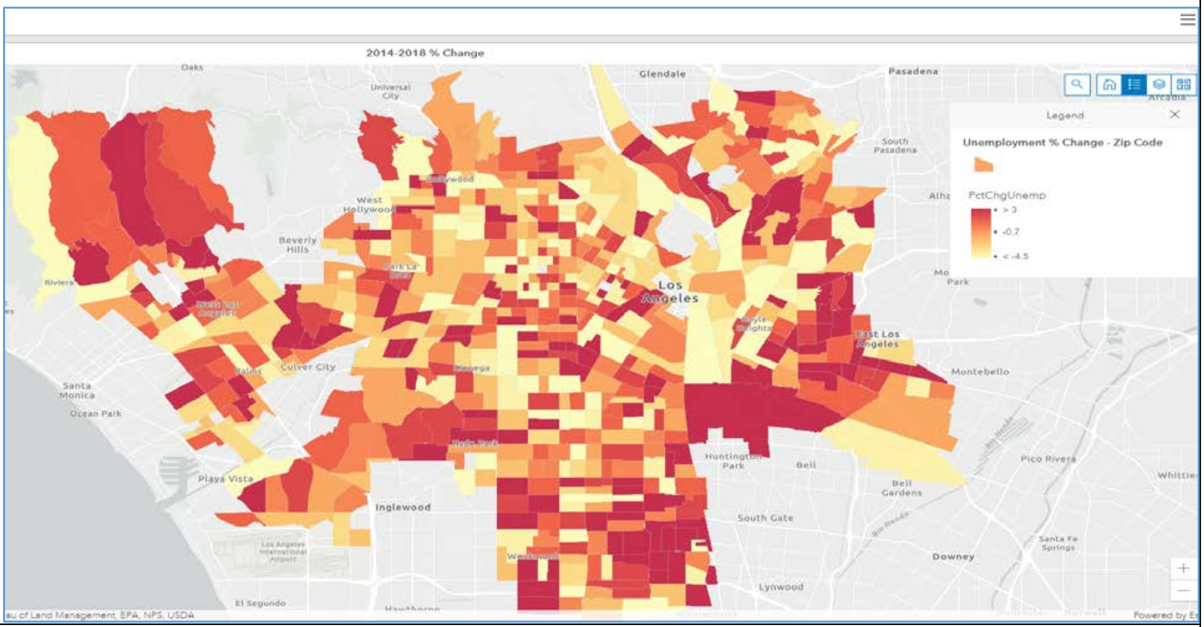


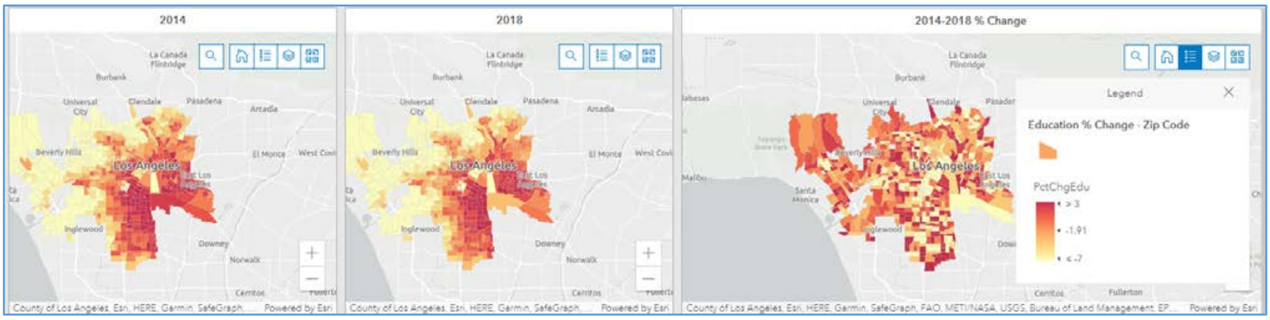
Figure 4c: City of LA – Indicator: Unemployment. (2014-2018 % Change)



Unemployment: Percent of the population over the age of 16 that is unemployed and eligible for the labor force.

Spatial intelligence garnered from these maps is indicative of the fact that from 2014 to 2018 we see unemployment staying consistent in the southern part of the City of LA. We do see improvements in the north eastern areas of the City. The City's south east and north west parts are candidates for improvement.

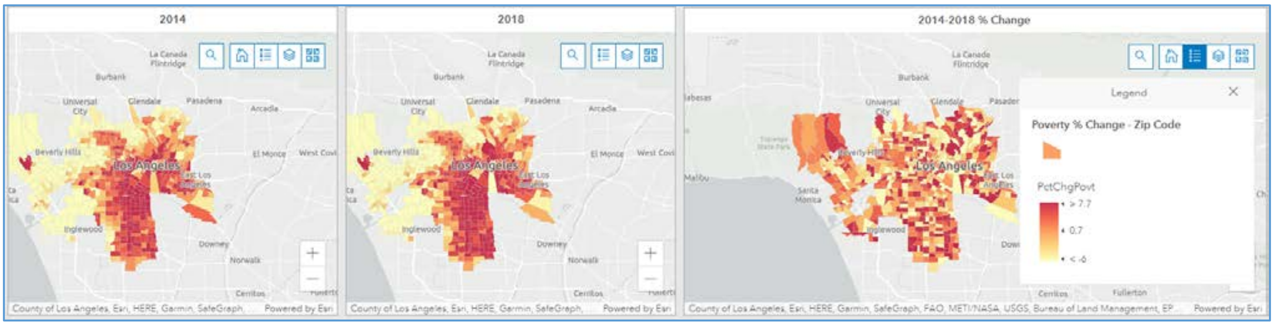
Figure 4d: City of LA – Indicator: Education. (2014, 2018, % change 2014-2018)



Education: Percent of population over 25 with less than a high school education

Spatial intelligence garnered from these maps is indicative of the fact that from 2014 to 2018 we see more people in the north western city portion with less education. It seems to be widespread rather than just concentrated in the southern and eastern parts of the City of LA.

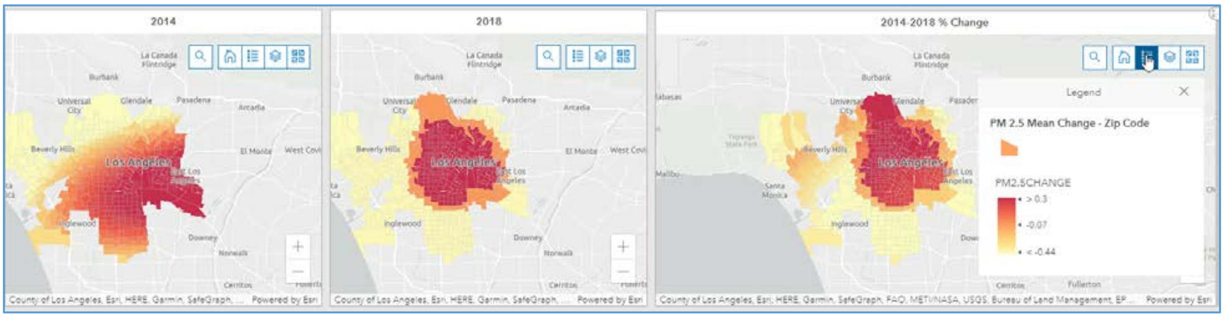
Figure 4e: City of LA – Indicator: Poverty. (2014, 2018, % change 2014-2018)



Poverty: Percent of population living below two times the federal poverty level

Spatial intelligence garnered from these maps is indicative of the fact that from 2014 to 2018 poverty has been expanding spatially from being concentrated in the central and southern parts of LA to being distributed more widely in 2018. We see areas in the north west city portions changing over the four years to show higher poverty levels.

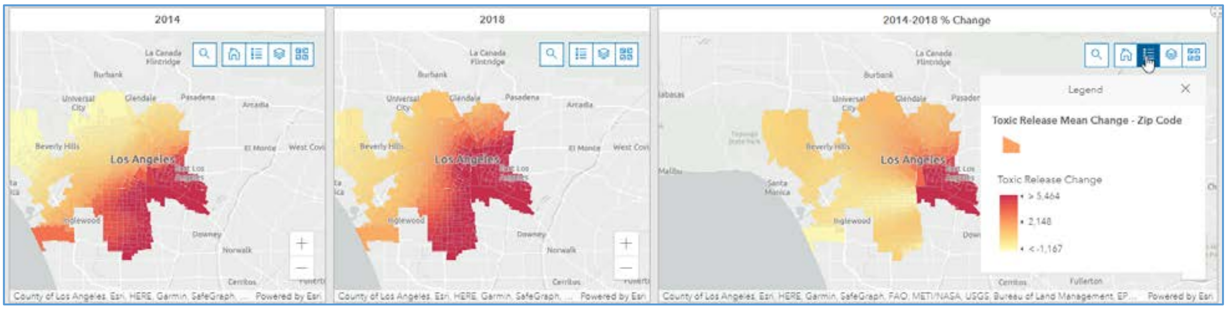
Figure 4f: City of LA – Indicator: Pollution. (2014, 2018, % change 2014-2018)



Pollution: Annual mean PM 2.5 concentrations

Spatial intelligence garnered from these maps indicates that from 2014 to 2018 the concentration consolidated in the central part of LA and the north city zone has worsened. However, the western part shows improvement from 2014 to 2018.

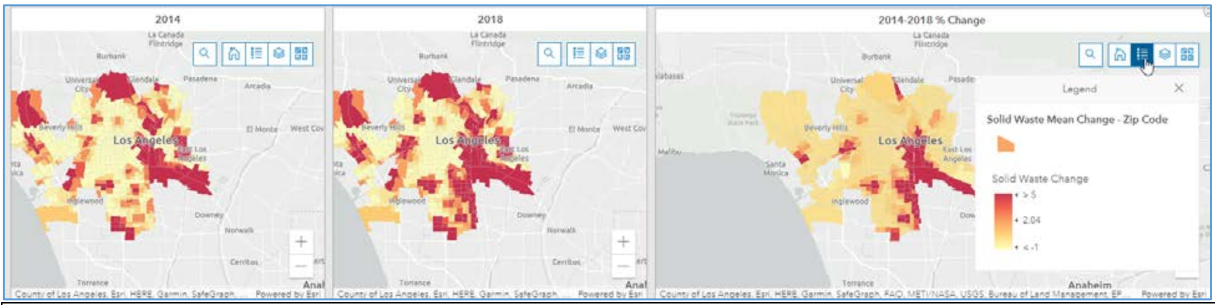
Figure 4g: City of LA – Indicator: Toxic Release. (2014, 2018, % change 2014-2018)



Toxic Release: Toxicity-weighted concentrations of modeled chemical releases to air from facility emissions and off-site incineration (from RSEI)

Spatial intelligence garnered from these maps is indicative of the fact that from 2014 to 2018 we see an improvement in the southern and eastern parts of the City of LA, but the south east continues to lag in improvements.

Figure 4h: City of LA – Indicator: Solid Waste. (2014, 2018, % change 2014-2018)



Solid Waste: Sum of weighted solid waste sites and facilities (SWIS) within buffered distances to populated blocks of census tracts

Spatial intelligence garnered from these maps indicates that from 2014 to 2018 we see an improvement in the northern and western parts of the city while the south east has had no visible improvement.

Studies have been conducted that examine spatial associations between different indicators. Voigtlander et al., (2010) research has shown that unemployment is indicative of neighborhood deprivation and that is associated with pollution exposure, leading to asthma and other health issues and increased susceptibility to toxic effects of pollution. Financial and emotional stress in residents of disadvantaged communities, often due to unemployment, also lead to poor health (Turner, 1995). Another study conducted (Roelofs et al., 2012) show that genetic changes are associated with exposure to the toxicity of solid waste emissions in landfill soil that contains a mixture of hazardous chemicals. This is known to be detrimental to the female reproductive system. We see that there are clear spatial autocorrelation effects visible from the descriptive maps. This has implications for the analysis of sustainable city success.

City planners, policymakers can use the maps and the underlying data to develop an analytical model that posits and examines the associations of sustainable city indicators to sustainable city success.

Benchmarking - Sustainability Changes 2014-2018 in City of Los Angeles

Building on the research framework (Figure 2) and utilizing DSR methodology for artifact development and refinement (Figure 3), the six indicators are mapped for sustainability for the City of LA by using GIS to gain location intelligence. The analysis conducted compares the City of LA's sustainability indicators over the period 2014-2018 using descriptive statistics and location analytics. We have added map layers to view the areas impacted with the change in 2018 compared to 2014.

The maps in Figures 4a to 4h clearly demonstrate the importance of location and the ease of analysis provided by longitudinal map visualizations.

Based on analysis of aggregate citywide totals, we believe that the City of LA has shown a slight improvement in indicators for unemployment, education, pollution PM2.5 and should put special focus on reducing poverty, toxic release and solid waste which has more than doubled in amounts (see Table 2) from 2014 to 2018. This is an opportunity for the City of LA to find solutions to improve performance with respect to indicators that are showing an undesirable increase over time for e.g. toxic release.

City of LA has ranked 2nd in the U.S. with a score of 34.5 points out of maximum of 55 for deployment of applications involving mobility, security, utilities, healthcare, economic development, housing and community (Woetzel et al., 2018 pg.12) and has partnered with University of Southern California and major tech companies and set up Intelligent IoT Integrator I3 consortium and has taken a leadership role in the sustainable city landscape. The city has been progressing since 2013 “from a moderate digital status to a leading digital city.” (Pick, 2016). They use IoT for boosting environmental awareness, livability, mobility, citizens’ safety, and lifestyle (Stone, 2018). As the City of LA is heavily invested towards sustainability applications, we see improvements over time in some sectors and in other sectors there is potential for improvement.

Table 2: Change in Sustainability Indicators (City of LA 2014-2018)

Indicators	**2014**	**2018**	**% Change 2014 - 2018**	**Summary of Results**
Social Indicator: Unemployment as % of population over age 16	12.02	11.31	-6.00	Overall Improvement. Unemployment for population over age 16 has gone down in 2018 from 2014. E.g. zip codes: 90049 (Brentwood neighborhood), 90042 (Highland Park neighborhood). There are zip codes that need improvements: 90058 (borders of city of Vernon), 90007 (near Exposition Park/University Park).
Social Indicator: Education: % of pop. over 25 with less that high school education	32.97	30.93	-6.00	Overall Improvement. Population over 25 with less than a high school education in 2018 has gone down from 2014. E.g. zip codes: 90058 (border of city of Vernon), 90040 (border of city of Commerce). There are zip codes that need improvements: 90005 (central LA), 90037 (central LA).
Economic Indicator: % of Population Living below Twice the Federal Poverty Level	51.50	52.24	-1.00	Needs improvement. Population living below two times the federal poverty level has gone up slightly in 2018 from 2014. E.g. zip codes: 90063 (east LA), 90012 (Chinatown area). There are zip codes that show an improvement:

				90041 (north east LA), 90031 north east LA).
Environmental Indicator: Air Pollution in Particulate Matter (PM) Annual, Mean	12.52	12.50	-0.02	Improvement. Annual mean PM 2.5 concentration has gone slightly down in 2018 from 2014. E.g. zip codes: 90019 (mid-Wilshire neighborhood), 90034 (neighborhood of Palms). these zip codes need improvements: 90068 (Hollywood Hills neighborhood), 90027 (Griffith Park).
Environmental Indicator: Toxicity Weighted Concentrations of Chemical Releases to Air, Mean	2.89	4.53	57.00	Needs improvement. Toxic Release has gone up in 2018 from 2014. E.g. zip codes: 90023 (Boyle Heights), 90022 (east LA). While these zip codes show improvement: 90045 (Playa Del Rey), 90001 (south LA).
Environmental Indicator: Sum of Weighted (ton) Solid Waste from Sites and Facilities near Population, Mean	1.09	2.31	112.00	Needs improvement. Solid waste has more than doubled in 2018 from 2014. E.g. zip codes: 90021 (downtown LA), 90002 (southeast LA). While these zip codes show improvement:90044 (south LA), 90049 (Brentwood).

The results of the longitudinal mapping and benchmarking of consolidated city indicators over time to measure success for the City of LA suggest that our proposed design principles are effective. Based on our findings, we propose policy development to improve sustainability in particular zip codes highlighted as trending towards lowered sustainability over time (see Table 2, right column).

DISCUSSION, IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH

Discussion

The objective of our study was to design a framework for a sustainable city that includes methodology, artifacts, sustainability goals and indicators, and relevant theories. The two research questions focus on location analytics as a tool for decision support and identification of which indicators in the City of LA has been successful over time and which ones need attention and using Design Science framework to produce artifacts to analyze city sustainability, frameworks and benchmarking.

Based on the findings, our objective is supported by our research framework in Figure 2. We believe it is valuable to follow this framework as it outlines steps that will be useful to identify SDG targets from high-level targets to extremely focused indicators to pinpoint the relevant indicators that are important at state (GPI) and city (mayor) levels. We have identified TOE theory, the DSR methodology and the resultant artifacts that are crucial for the success of a city desiring to be sustainable. In Figure 1 we have identified elements of T-O-E theory will be beneficial for technology adoption by cities since the drive to sustainable cities is dependent on ICT's. By identifying the city level elements instrumental for the technological innovation decision making impacted by the organization, the environment and the technology elements, we have provided a guidepost to cities desiring success in their SDG targets. Cities can also use these to benchmark to see at any given point where they are in comparison to other sustainable cities.

In Figure 3, we identified DSR as a methodology that will assist a city to be methodical in their approach towards the instantiation of a sustainable city. The iterative process of artifact development is crucial. The artifacts have been iteratively fine-tuned such as the location analytics maps that will assist a city to plan investments for their future growth wisely, and the frameworks that have been iteratively developed to identify elements relevant to a city's sustainability success.

Our location analytics in Figure 4a to 4h show the visual and numeric changes from 2014 to 2018. This is relevant to our decision making since a visual is a powerful aid for decision making. In Table 2 summarizing the sustainability indicators and analyzing by the % change allows us a tabular way to display our analysis and identify areas in detail that are successful or need further intervention to meet sustainability goals/targets.

With our research and focus on City of Los Angeles as a city that has invested in location analytics for decision making and is focused on sustainability, we have successfully provided frameworks and identified applicable elements of theory, provided artifacts that have been iteratively developed and suggested visual tools to pinpoint areas for sustainability. By following our methodology, a city can be successful to meet the UN SDGs relevant to their sustainability goals. Challenges such as higher rate of energy consumption, decrease in public health, increase in traffic congestion, worsening air and water pollution, toxic waste disposal, resource depletion, and social inequality and vulnerability can be met by following the study's methodology. In spite of unplanned urbanization, investment in Technology (ICT) can help mitigate social, environmental, and economic deterioration impacts and allow cities to use these technologies to achieve sustainability, reliability, energy efficiency, and productivity by applying integrated approaches to build institutional capacity and achieve their sustainability goals.

Theoretical Implications

DSR is a problem-solving paradigm, in which iterative design cycles help generate artifacts that are improved upon by continuous and incremental refining of the artifact. Incremental improvement in the artifact leads to gaining practical knowledge. We followed the relevant, rigor, and design/evaluate cycle of DSR methodology. In the relevance cycle, we identify indicators for the City of LA, examine the environment namely the city, residents, businesses and government entities to identify the indicators needed to meet the UN SDG's, and recognize the possible opportunity available to them upon problem resolution which is to make the city sustainable by using ICT. Among other things, a city can leverage people who have prior experience and are experts in this field to assist with progressing to a more sustainable city. In the rigor cycle, the existing knowledgebase was studied, and T-O-E theory was tested for applicability to ICT adoption. In the design and evaluate cycle, the systems development lifecycle phases are utilized to assess and refine the design.

We worked through an iterative process for the longitudinal location analytics mapping to attain visualizations of indicators for the City of LA appropriate for our study. We worked with various color schemes and icons, data sets and map layers for different time periods to assess spatial change over time and pinpoint areas for improvement. Location analytics helps to harness the power of geospatial data to uncover spatial patterns and trends, examine spatial relationships between key variables, and provides insights for decisions and actions. The iterative process has been instrumental in successfully delivering and displaying appropriate color scheme, icons, and layers for the location analytics maps artifact. By applying DSR methodology, cities have a structured approach to evaluating their environment for the problem they are trying to solve, use existing knowledgebase as well as design artifacts that have been iteratively improved upon to assess and evaluate their approach to meet the UN SDG targets.

DSR artifacts in this empirical study that answers our RQ2 are:

- Location Analytics Maps
- Frameworks:
    - Framework for T-O-E
    - Research Framework
- Instantiation Methodology

Limitations and Future Research Direction

Due to time limitation we focused on only four UN SDG targets. We picked City of LA as an example since the city is involved with location analytics and has sustainability targets. Additionally, data was available for those targets to analyze change over time. As part of future research, further analysis can be done for each goal to pinpoint precisely the various factors not covered in this paper that could potentially influence sustainability in a given area.

This paper restricts itself to the City of LA and six indicators for sustainable city success. Future studies can focus on a broader set of indicators and include a longitudinal sample of other sustainable cities for comparison. Additionally, layers for population growth can also be considered to allow cities to plan their infrastructure support for ongoing urbanization. We have used descriptive observational learning. Future studies can be conducted using smart city applications and machine learning to analyze how city spatial distributions change over time.

CONCLUSION

Research question 1 is supported by showing the relevancy of using location analytics to gain location intelligence. This technology is powered by GIS and rooted in science. With GIS, data is managed, visualized and analyzed to gain insights by asking questions about the data's surrounding and summarizing themes, spatial statistics and analytics on a visual map. We are able to identify hidden relationships and patterns by observing data and uncovering opportunities for improvements. This particular artifact is extremely useful to visualize and highlight areas in the city that are successful or need more attention. Location analytics assists with monitoring, assessing and course correcting as cities go along their journey to meet the UN 2030 agenda. For our study, we used location analytics to visualize and analyze data to assess change and uncover opportunities for improvements. We mapped the sustainability variables over the period 2014 – 2018 and showed spatiotemporal changes in key variables across the City of LA. This allows the city to identify areas they are successful in or need to pay careful attention to for further improvement.

Research question 2 is supported by showing the importance of design science frameworks to produce artifacts that have been refined by iterative development methods. The artifacts from this empirical study are: Location Analytics maps, T-O-E framework, research framework, and instantiation methodology. These artifacts are effective to guide cities to be successful in meeting the UN SDG's. The iterative development of maps has been instrumental in successfully delivering and displaying appropriate color schemes, icons and layers for the artifact. The frameworks for Research and DSR methodology allow cities to benchmark and use an established framework that has proved useful to the study of City of LA's sustainability goals and identifying areas that are successful or are a candidate for intervention. By using DSR as a problem-solving paradigm, a city can follow an established methodology that allows for opportunities to resolve a problem, understand the stakeholders impacted, and technology that can be implemented as a solution to mitigate the problem at the same time it is informed by theory that is applicable for the city to become successful with their sustainability goals.

We think this research is important and contributes to the MIS body of knowledge as our paper has proposed a research framework for sustainability, based on applicable elements of T-O-E theory and DSR methodology that include the location analytics mapping as an improved artifact to analyze sustainable city success in meeting UN SDG targets. Our research helps identify and analyze economic/social/environmental indicators and instantiates mapping the change in the economic/social/environmental indicators over time and space for the city. It also allows cities to benchmark and find ways and means to reach the targets as laid out in the UN SDG Agenda 2030 and become sustainable to attract investments. The analysis allows city officials to take corrective action where needed thus ensuring they can successfully meet UN SDGs.

Cities need to be sustainable to support urbanization and encourage profitable business investments. Meeting UN SDGs targets while considering indicators that are important at state and city level and assessing and evaluating them using location analytics is critical. Using frameworks and DSR methodologies as repeatable processes will increase the likelihood of successful delivery of new and improved sustainability projects.

Cities can benchmark and aim to match or better other sustainable cities. We identified elements from T-O-E theory that are applicable to a city in its adoption and benchmarking of technology. A city can benchmark and adopt elements from this theory for technology adoption and influence design of government policies and achieve the desired sustainability outcomes.

Our paper provides a research framework that a city can follow to become sustainable and to meet SDG targets. The present research framework helps to identify sustainability targets, apply elements from theory for technology adoption as becoming sustainable is dependent on ICT's so technology adoption becomes essential for success.

ACKNOWLEDGEMENT

The authors would like to acknowledge Dynasty Machado, University of Redlands alum for her contribution towards map design and generation.

REFERENCES

Amini, M. H., Arasteh, H., Siano, P. (2019) "Sustainable Interdependent Networks II", Studies in Systems, Decision and Control Springer Nature Switzerland AG. 186. https://doi.org/10.1007/978-3-319-98923-5_3. Accessed 2/26/21

Bagstad, K., Berik, G., Gaddis, E., (2014). "Methodological developments in US state-level Genuine Progress Indicators: Toward GPI 2.0. Ecological Indicators". 45. 474-485. 10.1016/j.ecolind.2014.05.005.

Baker, J. 2011. The Technology-Organization-Environment Framework, in In Dwivedi, Y. K; Wade, M. R.; Schneberger, S. L. (eds.). *Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1*. Integrated series in information systems. **28**. New York: Springer. pp. 231–245.

Bibri, S., and Krogstie, J. (2017). "Smart Sustainable Cities of the Future: An Extensive Interdisciplinary Literature Review". Sustainable Cities and Society. 31. 10.1016/j.scs.2017.02.016.

California Office of Environmental Health Hazard Assessment (2014) https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-version-20. Accessed 1/20/21

Ibid. (2018) https://oehha.ca.gov/calenviroscreen/maps-data/download-data. Accessed 1/20/21

Chatterjee, S., (2015). "Design Science in Information Systems Research" Lecture on 9 August 2015, Claremont Graduate University, Claremont, CA. Class Lecture.

Daly, H.E., Cobb, J., (1989). "For the Common Good: Redirecting the Economy Towards Community, the Environment, and a Sustainable Future". Beacon Press, Boston. Dietz, S., Neumayer, E., 2006.

Esri (2013). "The Four Imperatives of Location Analytics". https://www.esri.com/about/newsroom/arcuser/the-four-imperatives-of-location-analytics/. Accessed 2/13/21.

Esri. (2021) "Location Intelligence-Driving Digital Transformation" https://www.esri.com/en-us/location-intelligence?adumkts=product&aduse=&aduin=&aduc=print&adum=article&utm_Source=print&aduca=arcgis_promotions_awareness&aduco=arcnews-spring-2020-show-flier&aduat=article&adupt=awareness&sf_id=7015x000000iNpfAAE Accessed 2/26/21

Fox, M-J.V., and Erickson, J. D., (2018). "Genuine Economic Progress in the United States: A Fifty State Study and Comparative Assessment," Ecological Economics, Elsevier, vol. 147(C), pages 29-35.

Garcetti, E., (2021). “Sustainability”. https://www.lamayor.org/sustainability. Accessed 9/13/21.

Hevner, A., and Chatterjee, S., (2010). “Design Research in Information Systems: Theory and Practice”. Springer, May 2010. pp.9-22

Hevner, A. R., March, S. T., Park, J., Ram, S., (2004). “Design Science in Information Systems Research”, MIS-Quarterly, Vol.28, No.1, pp.75-105.

Hoornweg, Daniel A. (1961). “Cities and sustainability: a new approach”, Abingdon, Oxon; New York, NY: Routledge, 2016.

IMD and Smart City Observatory 2020. “Smart City Index 2020”. https://www.imd.org/smart-city-observatory/smart-city-index/

Lippert, S.K., Govindarajulu, C., (2006) "Technological, Organizational, and Environmental Antecedents to Web Services Adoption," Communications of the IIMA: Vol.6: Iss.1, Article 14. https://scholarworks.lib.csusb.edu/ciima/vol6/iss1/14

Mayor’s Office Los Angeles, (2019). “LA’s Green New Deal Sustainable City Plan, 2019”

Nidumolu, R., Prahalad, C.K., Rangaswami, M.R. (2009). “Why Sustainability is Now the Key Driver of Innovation” Harvard Business review, pp.57-64

OECD (2020). “Smart Cities and Inclusive Growth”, Ministry of Land, Infrastructure and Transport, Korea.

Pick, J., (2017) “Smart cities in the United States and worldwide: A rich arena for MIS studies”, Journal of Information Technology Case and Application Research, 19:3, 133-144, DOI: 10.1080/15228053.2017.1363597

Roelofs D., de Boer M., Agamennone V., Bouchier P., Legler J., van Straalen N. (2012). “Functional environmental genomics of a municipal landfill soil”. *Front Genet* **3**:85.

Stone, M., (2018) “The City of Los Angeles Is Taking Sustainability Data to New Heights”. https://www.delltechnologies.com/en-us/perspectives/the-city-of-los-angeles-is-taking-sustainability-data-to-new-heights/ Accessed 4/29/2021.

The United States Conference of Mayors (2018). “Cities of the 21st Century – 2018 Smart Cities Survey”. https://www.usmayors.org/wp-content/uploads/2018/06/2018-Smart-Cities-Report.pdf. Accessed 2/13/21.

Tornatzky, L. & Fleischer, M. (1990) The process of technology innovation, Lexington, MA, Lexington Books.

Turner J. B. (1995). “Economic Context and the Health Effects of Unemployment”. Journal of Health and Social Behavior 36(3):213-29.

UN (2017), Department of Economic and Social Affairs, Population Division (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP/248. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf. Accessed: 1/1/2018.

UN (2014), Department of Economic and Social Affairs, Population Division (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Highlights.pdf. Accessed: 1/1/2018.

UN (2015).” Transforming Our World: The 2030 Agenda for Sustainable Development. United Nations.” (A/RES/70/1). https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf. Accessed 2/1/21.

US Census. (2021). “American Community Survey”. Washington, DC: US Bureau of the Census.

USC Vietrbi (2021). “I3 Consortium”. https://i3.usc.edu/ABOUT/I3-CONSORTIUM/ Accessed 4/29/2021.

Voigtlander S., Berger U., Razum O. (2010). “The impact of regional and neighbourhood deprivation on physical health in Germany: a multilevel study”. BMC Public Health 10:403.

WCDE (World Commission on Environment and Development), (1987). “Our Common Future”. Oxford University Press, Oxford

Wieringa, R. J. (2014). “Design science methodology for information systems and software engineering”. Springer. https://doi.org/10.1007/978-3-662-43839-8

Woetzel, J., Remes, J., Boland, B., Lv, K., Sinha, S., Strube, G., Means, J., Law, J., Cadena, A., von der Tann, V. (2018). “Smart Cities: Digital Solutions for a more Livable Future”, McKinsey Global Institute.

DECISION SCIENCES INSTITUTE

Design and Analysis of the Smart Universal Onboard Truck Weighing System for Fuel Consumption Optimization Model

Jianyuan Peng
Georgia Institute of Technology
Email: pjy@gatech.edu

Shu Wang
Georgia Institute of Technology
Email: shu.wang@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

In this paper, the hardware system of the smart onboard truck weighting system was proposed. The frequency of the suspension is detected via two acceleration sensors mounted on the frame and the axle of the truck. The fast Fourier transform is used to convert the frequency data to the truck's weight. Then the data, including weight and location information, will be sent to the cloud-based dispatching system, which will implement the Fuel Consumption Model for Capacitated Vehicle Routing Problem (FCVRP) to do the real-time scheduling and fleet operation optimization.

KEYWORDS: Smart Sensing, Routing Problem, Fleet Operation Planning, Optimization, and Real-Time Scheduling

INTRODUCTION

Freight transport by trucks is one of the most common methods of transporting goods around the world. In addition, Covid-19 made people more willing to shop online, and services like one-day delivery are much more popular than before. However, doesn't like train, ship, or aircraft, the transporting conditions for trucks are much more complex than the rest methods. The trucks are sharing the same road with passenger cars and other commercial vehicles so that more factors need to be considered to deliver goods on time than the rest of the methods. To solve such problem, a better fleet management strategy is necessary.

One of the most critical aspects of fleet management is route planning for each truck. There are multiple destinations for the whole fleet each day, with different distances, road conditions, and traffic conditions. The range that a track can go will be determined by those factors in addition to the truck's fuel consumption rating, driver's driving habit, and cargo weight. The last factor is important, especially to the vehicles powered by new alternative energy vehicles like electric trucks.

This paper will be focused on the truck's weight. It will be detected via sensors and converted to useful data and enable a better route planning strategy. There are three technical challenges about the weight detecting system and the following fleet management strategy. Firstly, the sensor kit must be universal, easy to install, and does not affect any safety factors of the truck. Secondly, the raw data collected from the sensor must be converted to useful and accurate data. And thirdly, those extra data musts be useful to make correct decisions that improve the transporting efficiency and reduce costs.

To achieve first two challenges, the truck suspension was simplified as a 1-degree-of-freedom mass-spring-damper model. A couple of acceleration sensors were put on the axle and the frame of the truck. By doing fast Fourier transform and fitting the theoretical model, the weight of the truck can be detected accurately. To achieve the last challenge, a fuel consumption model was implemented for the classic Capacitated Vehicle Routing Problem.

RELATED WORK

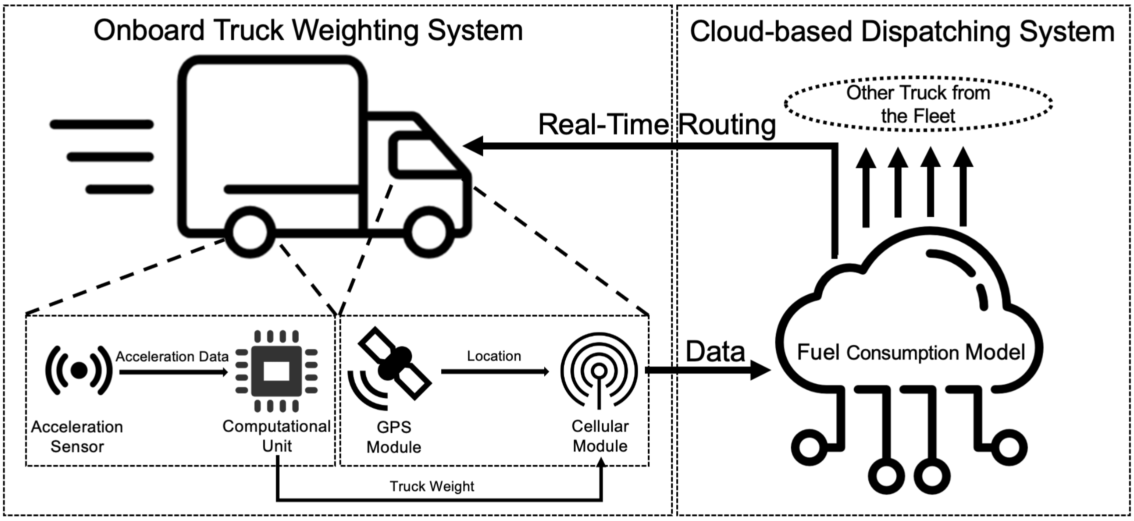
There are various existing methods to detect the weight of the truck. The most obvious way is to periodically put the truck on a weight scale. But the weight scale is not always accessible, and the weight data will not be real-time and continuous. There are other types of onboard sensors, such as measuring the pressure inside the vehicle's suspension airbag (Right Weigh, 2018), measuring the stress within the leaf spring (Chen, 2011), and measuring the length of the shock absorbers (Ronen, 1988). However, those methods are usually only suitable for one specific type of suspension. The installation is invasive and may affect the mechanical performance and safety of the truck, the price is costly, and no other modules to support the data transmission to the dispatching system.

The Vehicle Routing Problem (VRP) is an optimization problem about finding the optimal set of routes for the fleet in order to deliver goods to a set of customers (Dantzig & Ramser, 1959). More variants of the classic VRP are evolved like the Multi-Depot Vehicle Routing Problem (MDVRP) which multiple depots exist from the start point to the destination (Nagnyd & Haque, 2019), and Capacitated Vehicle Routing Problem (CVRP), which the carrying capacity of vehicles is limited. The CVRP with Fuel Consumption Rate (FCVRP) was recently discovered (Xiao et al, 2012), where the optimization object is to minimize the total fuel consumed instead of minimizing total distance traveled.

SYSTEM DESIGN

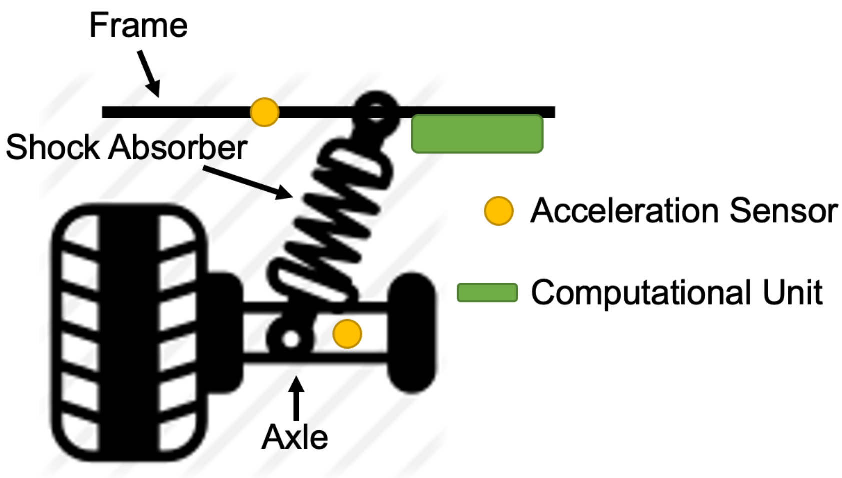
The overall system contains two parts: an onboard truck weighing system and a cloud-based fleet dispatching system, as shown in figure 1. The onboard truck weighing system includes two acceleration sensors, a computational unit, a GPS module, and a cellular module. The acceleration sensors will detect the acceleration of the truck frame and the axle, which will be used to calculating the real-time truck weight via fast Fourier transform done by the computational unit. Then combining with the location information from the GPS module, the cellular module will send all those data to the cloud-based fleet dispatching system. The dispatching system will analyze the data from the fleet and generate the real-time management of routing based on the fuel consumption optimization model and dispatch the job to the truck driver.

Figure 1: Overall Technical Framework of the Smart Universal Onboard Weighing System for Fuel Consumption Optimization Model



The onboard system is easy to install/maintain and will not affect any mechanical structure of the truck. The sensors and the computational unit will be installed under the car with double-sided tapes and zip ties. Figure 2 shows the possible location for those sub-components.

Figure 2: The installation of the Onboard System Sub-components

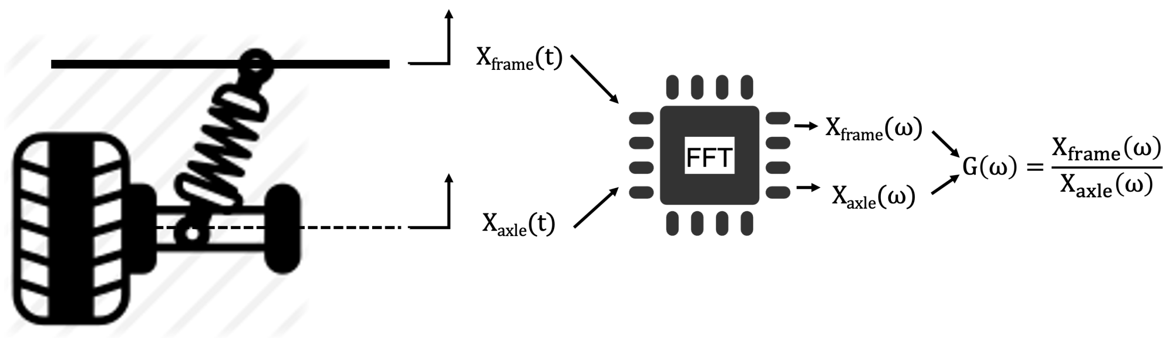


The GPS and cellular module will be installed inside the truck cabin. The data from the computational unit will be transferred wirelessly to the cellular module, and all components will be powered by the truck’s 12-volt battery.

MASS-SPRING-DAMPER MODEL FOR WEIGHT DETECTION

The dynamic model of the truck suspension is simplified to a ¼ vehicle suspension model of 1-degree-of-freedom. The simplified model is shown in figure 3.

Figure 3: Simplified ¼ Vehicle Suspension Model (1-DOF MSD System)



The transfer function of the 1st-order mass-spring-damper (MSD) system is

$$|\boldsymbol{H}(\boldsymbol{\omega})| = \left| \frac{\mathbf{1} + \frac{\mathbf{2}i\xi\omega}{\omega_n}}{\mathbf{1} - \left(\frac{\omega}{\omega_n}\right)^2 + \frac{\mathbf{2}i\zeta\omega}{\omega_n}} \right| \tag{1}$$

Where ω is the system frequency, ω_n is the suspension natural frequency and $\omega_n = \sqrt{\frac{k}{m}}$, ζ is the damping ration and $\zeta = \frac{c}{2\sqrt{mk}}$, m is the truck mass, k is the spring stiffness constant, and c is a constant.

The sensors will collect the data for the system frequency, and then a curve-fitting will be implemented for the above model to find the estimated natural frequency because the only parameter that is related to this application is the truck mass. Any other variables can be seen as truck-specific constants for the model. Thus, the empty truck mass should be known, and an initial calibration run is required before using this system on each truck. The relationship between the estimated natural frequency and empty truck mass will be used as a multiplier for the actual weight detection of the truck.

The preliminary testing was done on a small, 6350 lbs pick-up truck with leaf-spring suspension. Three conditions were tested that are empty truck, truck with 320 lbs of load, and truck with 1032 lbs of load. The test results are shown in figure 4 and table 1 with the maximum error of 6%. This is very accurate for the fuel consumption model that will be addressed in the next section.

Figure 4: Frequency Response and Calculated Mass of Different Loads

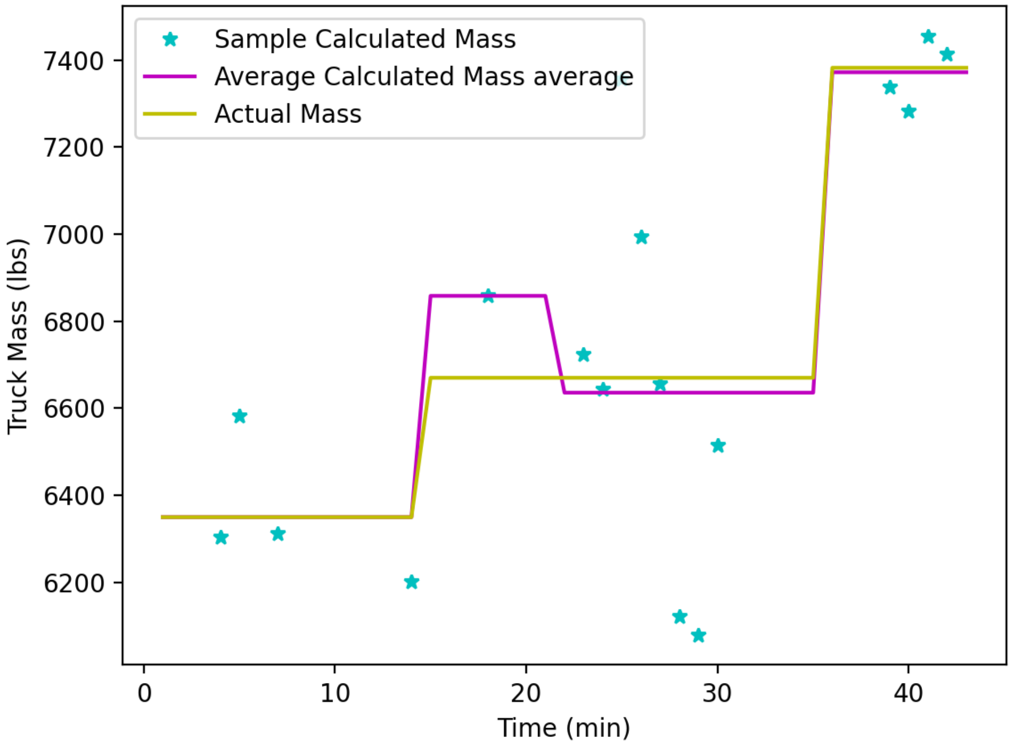


Table 1: Experiment Results and Error			
Test	Actual Mass (lbs)	Calculated Mass (lbs)	Error
Calibration	6350	-	-
1	6670	7066	6%
2	6670	6834	2%
3	7382	7450	1%

FUEL CONSUMPTION MODEL FOR CAPACITATED VEHICLE ROUTING PROBLEM

Statistics show that the distance a vehicle can travel per volume unit of fuel consumed is strongly correlated to the vehicle's weight. The weight information provided by the onboard weighing system can be directly used to calculate the fuel consumption rate of a vehicle. This section will discuss the fuel consumption rate formulation and the optimization model for vehicle routing problem considering fuel consumption rate.

Fuel Consumption Rate Formulation

Assuming the fuel price is $1.00/Liter. Figure 5 and table 2 shows the estimated fuel cost savings vs. different levels of weight reduction, according to Canada government website.

Figure 5: Estimated Fuel Cost Savings for a Range of Weight Reductions

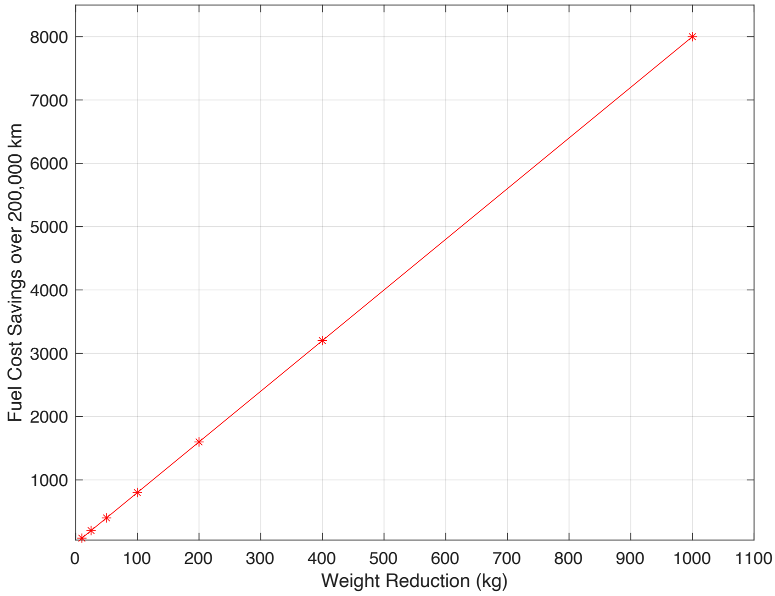


Table 2: Estimated Fuel Cost Savings for a Range of Weight Reductions	
Weight Reduction (kg)	Fuel Cost Savings over 200,000 km ($)
10	80
25	200
50	400
100	800
200	1600
400	3200
1,000	8000

Therefore, the fuel consumption rate can be simplified into a linear model

$$f(l) = a(m + l) + b \tag{2}$$

Where l is the load weight on the vehicle, m is the empty weight of the vehicle, and a, b are constants. Two initial calibrating run of the vehicle where there is no load and there is maximum load (l_{max}) should be done before implementing this model. So, assume the fuel consumption rate of those two extreme conditions are known

$$f_0 = am + b \tag{4}$$

$$f_{max} = a(m + l_{max}) + b \tag{5}$$

The constant b is irrelevant and the constant a can be written as

$$a = (f_{max} - f_0)/l_{max} \tag{6}$$

And then the fuel consumption model from equation (2) can be re-written as

$$f(l) = f_0 + al \tag{7}$$

Which is the simplified fuel consumption model in this application. With the fuel consumption rate, the fuel cost of one vehicle from node i to j is

$$Cost_{fuel}^{ij} = c_0 f_{ij} d_{ij} \tag{8}$$

Where c_0 is the unit cost of the fuel, f_{ij} is the fuel consumption rate from node i to j, and d_{ij} is the distance between node i and j.

Optimization Model for Vehicle Routing Problem Considering Fuel Consumption Rate

The objective function of the classic Capacitated Vehicle Routing Problem is to minimize the total distance travelled by all vehicles in the fleet shown in equation (9), assuming that there is n customers

$$Z_{CVRP} = min \sum_{i=0}^{n} \sum_{j=0}^{n} d_{ij} X_{ij} \tag{9}$$

Where X_{ij} is a binary variable shows if the vehicle k can travel from node i to j. Instead of minimizing the total travel distance, with the consideration of fuel consumption rate, the optimization model is to minimize total fuel cost and can be written as

$$Z_{FCVRP} = min \left(\sum_{j=1}^{n} F x_{0j} + \sum_{i=0}^{n} \sum_{j=0}^{n} c_0 d_{ij} (f_0 x_{ij} + a l_{ij}) \right) \tag{10}$$

With constraints of

$$\sum_{j=0}^{n} x_{ij} = 1, i = 1, \dots, n \tag{11}$$

$$\sum_{j=0}^{n} x_{ij} - \sum_{j=0}^{n} x_{ji} = 0, i = 1, \dots, n \tag{12}$$

$$\sum_{j=0, j \neq i}^{n} l_{ji} - \sum_{j=0, j \neq i}^{n} l_{ij} = D_i, i = 1, \dots, n \tag{13}$$

$$l_{ij} \leq l_{max} x_{ij}, i, j = 0, \dots n \tag{14}$$

Where F is the fixed cost of each vehicle, l_{ij} is the vehicle load from node i to j, and x_{ij} is a binary variable and equals 1 if node j is the next destination after node i, and otherwise equals to 0.

DISCUSSION AND CONCLUSIONS

In this paper, a hardware solution for smart universal onboard truck weighing system is proposed. The system is accurate, easy to install, and will not affect any mechanical structures of the truck. The most significant advantage of this system is that it can send real-time weight information to the cloud-based dispatching system, and no need to stop at a truck scale and manually send the “discrete weight data.” Then, the dispatching system can use this extra weight data to optimize the fleet schedule via solving Capacitated Vehicle Routing Problem with the consideration of fuel consumption rate (FCVRP) as the weight is strongly correlated to the fuel consumption rate. The objective function is then to minimize total fuel cost instead of distance traveled from the classic Capacitated Vehicle Routing Problem. This system will work for all sizes and weights of trucks theoretically. However, further experiments are needed to confirm this point with better calibrations and uncertainty analysis. The methods for solving the FCVRP are also needed to be implemented in further studies.

ACKNOWLEDGEMENTS

Bilal Ali, Matthew Moore, Aman More, Matthew Murray, George Smirnov, and Angie Zhang are acknowledged for their contribution of hardware prototype development and preliminary testing.

REFERENCES

Chen, Y. (2011). Intelligent vehicle load measuring system and method. U.S. Patent 7,881,903

Dantzig, G. B., & Ramser, J. H. (1959). *The Truck Dispatching Problem*. Management Science, 6(1).

Government of Canada. Vehicle Weight. Retrieved from https://www.nrcan.gc.ca/energy-efficiency/transportation-alternative-fuels/personal-vehicles/choosing-right-vehicle/tips-buying-fuel-efficient-vehicle/factors-affect-fuel-efficiency/vehicle-weight/21024, May 29.

Mahmud, N., & Haque, M. (2019). *Solving Multiple Depot Vehicle Routing Problem (MDVRP) using Genetic Algorithm*. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE).

Right Weigh. (2018). Air Suspension Load Scale. Retrieved from https://www.rwls.com/learn, May 29.

Ronen, D. (1988). *Perspectives on practical aspects of truck routing and scheduling*. European Journal of Operational Research, 35(2).

Xiao, Y., Zhao, Q., Kaku, I., & Xu, Y. (2011). *Development of a fuel consumption optimization model for the capacitated vehicle routing problem*. Computers & Operations Research.

DECISION SCIENCES INSTITUTE

Building transparency in supply chains using a layered approach

Dr Penina Orenstein
Seton Hall University
Stillman School of Business
Department of Computing and Decision Sciences
400 South Orange Ave
South Orange NJ 07079
Penina.orenstein@shu.edu

Petar Micevski
Seton Hall University
School of Arts and Sciences
Department of Mathematics and Computer Science
400 South Orange Ave
South Orange NJ 07079
petar.micevski@shu.edu

ABSTRACT

In earlier research, we presented the topological structure of supply chain networks in an aesthetic, visual form. But, aside from the structural information that emerged, it was difficult to gain any detailed insight about the supply network's operation. Essentially, we created a tiered visualization (wireframe) of the nodes and edges (firms) in the supply chain network.

In this research, we take the existing wireframe network and describe how it can be transformed into a dashboard. A dashboard is a form of data visualization which provides at-a-glance views of key performance indicators (KPIs) relevant to a particular objective or business process. We then identify several dimensions of information which should be included in the dashboard, such as structural, geographic as well as financial layers. The resulting dashboard tool is broad-based since, depending on the goal of the analyst, further layers may be applied to the structure. The information can then help improve the understanding of the supply network beyond confirming what is already known. This paper describes the information framework for the dashboard. The knowledge that is obtained through the analysis of the dashboard will yield areas for improvement and greater efficiency in the supply chain, reduce the chances of disruption and enable the supply chain to stay competitive.

KEYWORDS: Data Mining, Digital Visualization, Business Analytics

INTRODUCTION

Identifying and analyzing supply chains, from both financial and geographic perspectives, can help drive data-driven research by assessing the potential impact from events like Coronavirus, Zhu et al. 2020, geopolitical trade tensions, and more. Recent developments surrounding the

Coronavirus, have increased the need to understand the mechanism that drives global supply chains. In this research, we consider how data visualization can be employed as a tool to explore the supply chain operation. Data visualization is the representation of data or information in a visual format, essentially forming a bridge between the data and its graphical representation. We envisage that the knowledge which emerges can be used to understand trends and patterns in the data which may not be obvious from the actual data set. The background for this idea comes from previous research in which we created graphical maps that essentially transformed supply chain financial relationship data extracted from Bloomberg and FACTSET, into visual images of the supply chain network, Orenstein, 2020. In addition to the visual component, we also provided structural metrics which helped explain the evolutionary nature of these networks. Unfortunately, this information could not be used to make decisions about the supply chain because our understanding of its operation was impaired, for example, we could not trace an item through the supply chain network or explore the impact of disruption since we did not have sufficient knowledge of the supply chain. This led to a follow-up analysis, in which the authors applied the visualization algorithm to approximately sixty supply networks across six industry sectors, over a ten-year period. The impact of topological structure on the financial performance of the focal supply network was then explored, Orenstein and Tang, 2021. We found that average degree (a parameter which indicates network connectivity) is negatively correlated with firm performance. However, while this result increases our understanding of the financial capability of the supply network relative to its topological structure, we were limited in obtaining any detailed insight about the supply network's operation. In particular, we were not able extract any geographical information about the supply chain, for example what path do parts take in traversing the supply chain. Essentially, the graphical images helped confirmed what we already knew, that the global supply chain is complex, and that, even with the data visualizations, our understanding of the supply chain operation did not improve substantially. In a certain sense, we had produced a collection of pretty pictures, but no actionable decisions could be made because of analyzing the underlying metrics.

Our aim in this paper is to improve our understanding of the supply network so that recommendations about the supply network's operation can be made. In this paper, we enhance an existing visual data map using a layered approach. We take an existing wireframe network and describe how it can be transformed into a dashboard. A dashboard is a form of data visualization which provides at-a-glance views of key performance indicators (KPIs) relevant to a particular objective or business process. As a first step, we highlighted the nodes in the supply chain network by tier so we could understand how the networks evolve by tier. But more importantly, we added targeted information into the data construct so that, the end-result is a comprehensive map detailing the linkages between tier-one, tier-two, and tier-three suppliers visualized on a geographical map. In this way, we could potentially use the visualizations to be able to trace individual parts to the exact site where they are manufactured. Ultimately, the mechanism could be used to help companies ensure sustainability practices from the production of raw materials through the finished goods, and also possibly allow companies to identify and anticipate vulnerabilities in their supply chain. More importantly, we expect that the layered visualizations can be applied to unlock predictive analytics capabilities and help supply chain managers act proactively. In addition, the visual mapping coupled with the layers of information will then improve the comprehension of the supply network operation and ultimately contribute to the decision-making process as to how improve its efficiency. This is the primary goal of this research: to improve the comprehension of the supply network operation through actionable, quantifiable items.

LITERATURE REVIEW

To date the scope of visualizing a supply chain network via digital tracing has mostly been restricted to first tier primary suppliers, International Labor Organization, 2019. As supply chains become more complex, tracking a vast number of suppliers to trace the origins of products, particularly the raw sources from upstream in the supply chain, remains a significant challenge, Cernansky, 2020. The fragmentation and global dispersion of supply chains across international borders may obstruct the visibility of certain suppliers, making some areas of a supply chain opaque, (Norton et al., 2014). Additionally, traceability requires investment in technologies and processes that track goods along the supply chain, and require coordination and information sharing across different actors in the supply chain. These efforts can be costly and require time and willingness from all the respective actors, some of whom may be reluctant to share information due to legitimate business concerns about competitiveness, Fair Labor Association, 2017.

The broad question which supplies chain visualization seeks to address is “do you know who is in your supply chain?” Clearly, not all supply chain visualizations are equal. There are several approaches that can be employed to produce a comprehensive supply chain mapping. Ideally, a supply chain map should include all the supplier sites as well as the sub-tier information. Collecting sub-tier data is more difficult due to its proprietary nature. The ability to drill down in the supply chain will depend on the extent to which the information can be obtained.

Consider this anecdotal evidence provided by Meiklejohn, 2015, who works for Eileen Fisher as a Supply Chain Transparency specialist. Through this example, we can understand the need for a comprehensive supply chain mapping. She traces clothes from field to factory. In her case, a mapping is required to ascertain that the cotton used by the company is organic, and that the fabrics are safely dyed. But there is a significant by-product which arises in her role, as she collects data by speaking to suppliers along all points in the supply chain. Data from her conversations is collected and mapped to produce a digitally transparent, traceable map of the supply chain. She explains that inefficiencies in the supply chain can be highlighted through the mapping being created. For example, she speaks to manufacturers along the entire path of garment manufacturing, yarn mills, fiber manufacturers and the like, and identifies the source of all material along the path from manufacturer to customer. This is important because the company needs to know where the material is coming from to ascertain that it is entirely organic. If fiber crosses the globe to be spun, then traverses back to the US to be woven, is subsequently dyed across the sea, this is an inefficiency. Can the supply chain be re-engineered to make improvements and lower the carbon footprint? This is a key aspect of traceability.

Using the age-old adage, “you cannot manage what you cannot see”, the goal of the mapping/tracing in this study is to increase the level of visibility in the supply chain, in order to be able to trace products from source to destination and use this knowledge to improve its performance and make suggestions about re-engineering the supply chain to make it more efficient, less prone to disruption, and increase sustainability. These are all areas which have not been addressed until now, Frost and Sullivan, 2019.

The time is ripe for developing a mechanism to effect these extensive supply chain tracings, Vakil 2020. With the impact of Covid-19, reliance on Chinese suppliers has been identified as a critical weakness in manufacturing supply chains. The time has come for companies to focus on building a detailed supply chain mapping (trace), not only of first-tier suppliers, but also of second and third tier suppliers. The tracing will allow companies to track goods from source to destination as well as identify which countries form part of the path. It appears, that companies that invested in this level of visibility appeared to be in better shape after the pandemic hit.

In a follow-up article, Vakil 2021, expands on this point and highlights that when the pandemic hit, "70% of organizations did not have a sense of what parts of their supplier networks were affected." A large portion of these companies were focused on data collection and assessment mode trying to identify which local suppliers in China were impacted by the lockdown. By contrast, companies that had invested in supply chain risk management tools, particularly supply chain mapping appeared to have a different experience." They could conduct "what-if" analyses and work with suppliers to protect supply lines. Still, only a minority of companies used Fukushima and the Thailand floods as a wake-up call to gain visibility into their supply chain; a critical mistake when COVID struck. Those that set up comprehensive, multi-tier supplier mapping programs came into 2020 more prepared. According to Choi et al, 2021, "by having visibility into their supplier networks, companies such as GM, Cisco, IBM, and Amgen were able to quickly ascertain what parts and materials originated in Wuhan and Hubei and fast-track their responses".

It turns out that to map a supply chain one can obtain an overarching picture of the supply chain, but to really trace things properly and understand where the product traverses geographically, one needs to collect the data about the product. How it is manufactured, who are the suppliers? That is a lot more complex and requires attention to detail.

From this discussion, it is clear that the focus needs to shift towards creating transparent supply chains. The contribution of this paper is to identify the data sources that should be included in the mapping to create a digital tracing framework for supply networks.

DATA SOURCE

We assume that we can represent a supply chain network as an unweighted directed graph *G=(N,L)* with *N* nodes and *L* links or edges. We assume there are no cycles in the graph. In the context of this work, the nodes represent individual companies, which connect through edges. We use only supplier relationships. i.e., the degrees are all one directional. In the paper, we refer to *G* as the focal company, or focal node. Each edge represents a financial relationship between a pair of companies, essentially a binary relationship. As such, we can denote the network as an adjacency matrix *(A)*. Any element of the adjacency matrix $A=a_{ij}$, is given as:

$$a_{ij} = \begin{cases} 1, if\ i \neq j\ and\ i\ and\ j\ nodes\ are\ connected\ by\ an\ edge \\ 0, if\ i \neq j, and\ i\ and\ j\ nodes\ are\ not\ connected \\ 0, if\ i = j \end{cases} \tag{1}$$

To create a dynamic visualization map of a particular supply chain network, we first acquired financial relationship data for a prototype company and organized the data by calendar quarter and tier for a period of ten consecutive years (2005-2015). In a supply network structure, it is not uncommon for a given company to appear in different tiers. Each time the said company appears in a particular tier, it is recorded. Aside from the prototype company described in this paper, we have separately collected financial relationship data for approximately sixty companies that span various industry sectors including Energy, Health Care, Industrial, Communication and Information Technology for the ten-year time span. The sector classification is derived from Yahoo Finance (https://finance.yahoo.com/industries/), which extracts the information from the

GICS classification system (https://www.msci.com/gics). The data set that has been created is extensive. The goal of this research is to develop a dashboard structure, essentially a proof of concept. Hence, we choose to spotlight a single company from the healthcare sector: AMGEN. We also chose AMGEN since it was highlighted by Vakil et al, 2021, as a company with traceability capabilities. With the information displayed on the dashboard, we can then apply the technique to any network in our collection, Orenstein and Tang, 2021.

The AMGEN Data Set

AMGEN (Applied Molecular Genetics Inc.) is one of the world's largest independent biotechnology companies. Amgen was established in Thousand Oaks, California, in 1980. Amgen's Thousand Oaks staff in 2017 numbered 5,125 (7.5% of total city employment) and included hundreds of scientists, making Amgen the largest employer in Ventura County. Focused on molecular biology and biochemistry, its goal is to provide a healthcare business based on recombinant DNA technology, (Wikipedia, retrieved April 27th 2021).

Before analyzing the AMGEN network, we examine some financial data about the company. The information was obtained from Compustat. As can be seen in Figure 1, AMGEN has seen a remarkable growth in both their total assets as well as in their sales and market capitalization. The reports highlight 2013-2014 as a milestone year, in the sense that for the first time in AMGEN's history, the total company revenue surpassed $20 billion. From the graphs, one can observe that AMGEN's sales is on an upward growth curve contributing to strong financial health.

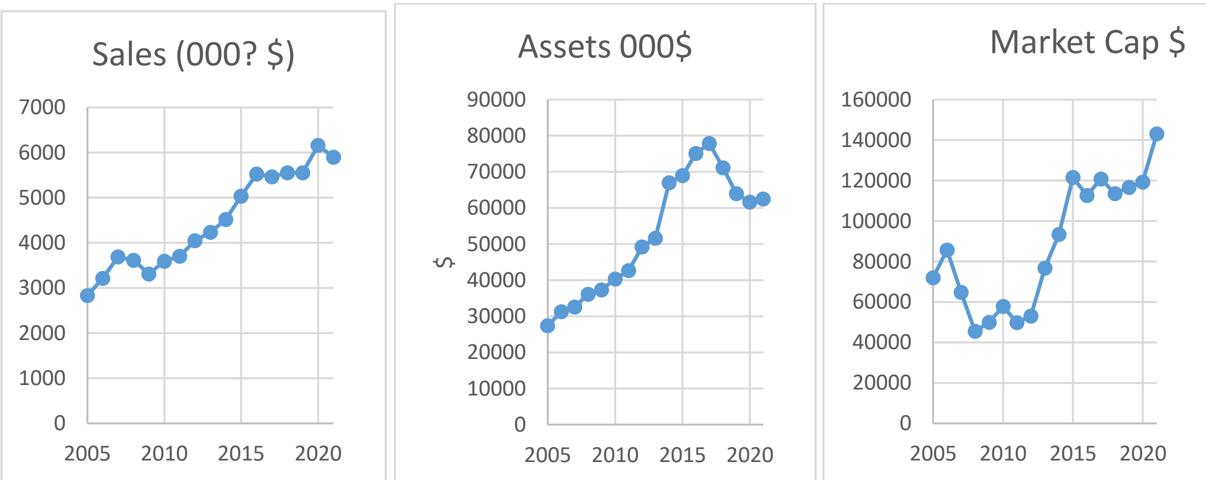


Figure 1: AMGN Total Sales, Total Assets and Market Cap: Source AMGN Compustat Data.

RESULTS

In this section we describe the framework which has been created to trace supply chain networks. The analysis starts with the basic network map and describes the stages leading to the detailed dashboard structure which can then be applied to help the analyst gain an understanding of the underlying supply chain operation. The application of the data dashboard and the results that can be obtained using the dashboard will be described separately, in a follow-up paper.

Creating the basic network map

The first stage of this analysis is to create a map of the network with no information other than the connections and the arcs that link the nodes together. An example of this is shown in Figure 2 which displays the AMGEN network in 2005, quarter 1. One can observe a sparse structure which spans from a core of nodes (hub) at the center of the network. The data source is FACTSET, but the imaging method may be applied to Bloomberg data in a similar fashion, (Orenstein, 2020).

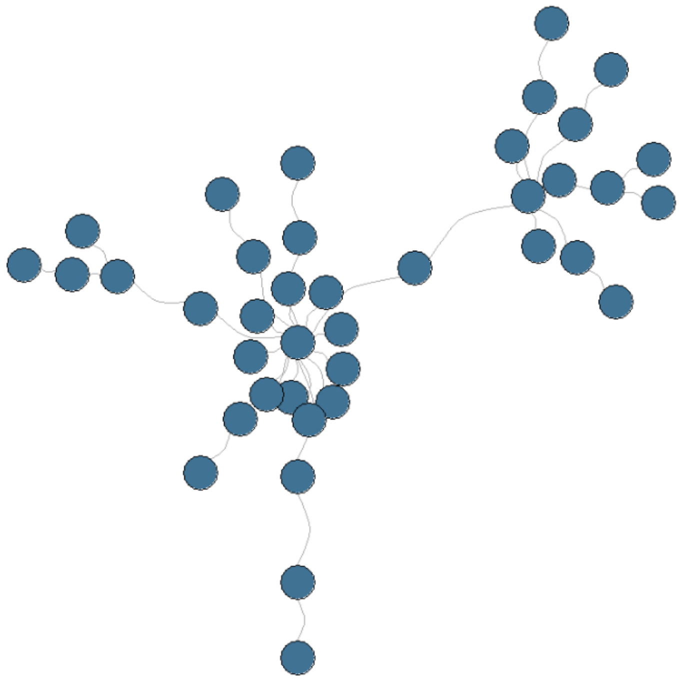


Figure 2: Topological form and visual map of the AMGEN network in 2005, Quarter 1; Source FACTSET

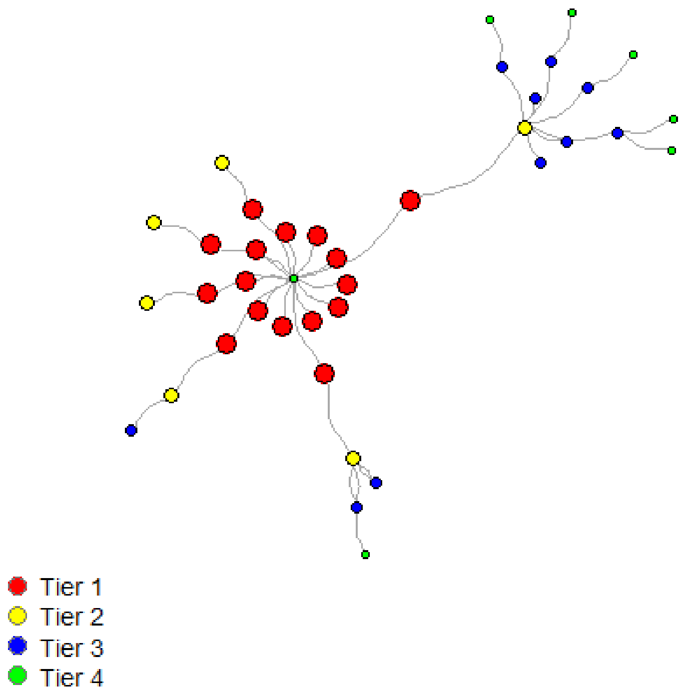


Figure 3: Visual map of the AMGEN network in 2005, Quarter 1, highlighting the tiered information ; Source: FACTSET.

Figure 3 expands on the initial mapping from Figure 2 because it highlights the nodes in the network by tier. One can examine the proportion of Tier 1 companies that comprise the network, and how the number of nodes within each tier changes as the network expands. The color indicates how many jumps to reach focal company.

One can also extend the mapping technique to compare how the supply network changes with time. This extension is applicable to FACTSET data only. The evolution of the network can be obtained by examining the historical financial relationship information by specifying the window of interest. Unfortunately, while Bloomberg data provides a detailed snapshot of the suppliers in the network at a particular point in time, one cannot look back to obtain a dynamic overview.

To explore the dynamic evolution we highlight Figure 4, which shows the AMGEN network in 2005, along with snapshots from 2009 and 2015. One can see how the structure of the core network evolves and note the expansion of the network at the upper tiers. There appears to be a correspondence between AMGEN's structural complexity and AMGEN's financial growth highlighted in Figure 1. A separate, comprehensive empirical study of the relationship between network structure and overall financial performance is under preparation, (Orenstein and Tang, 2022).

The next phase of our analysis takes the existing network structure (Figure 3) and applies additional layers of information to the visualization. We begin with three dimensions: financial, structural, and geographic. These dimensions, when superimposed on the existing network enable the user to gain further insight into the network's operation and explore beyond the pretty picture. The first step in this process is shown in Figure 5. One can see how a layer of financial data (firm size) depicted by the relative size of each company within the supply network. When a company within the supply network exhibits a significant revenue relative to the focal company, a label for that company is displayed. This enables the user to easily identify the key suppliers within the supply network. For example, in Figure 5(a), SAP, a Tier 2 supplier is a significant player in the AMGEN network. In subsequent Figures, (b)-(d) we can track the top suppliers and examine their influence.

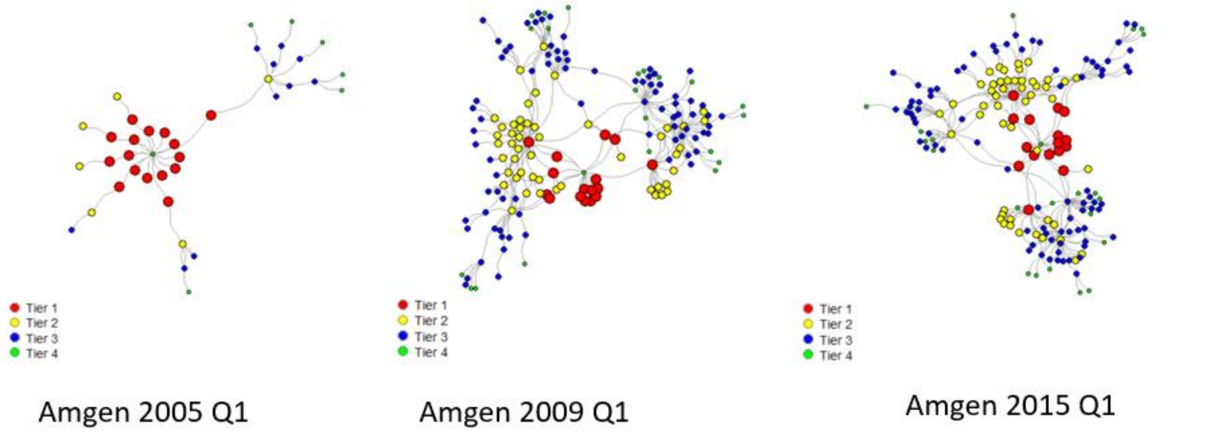


Figure 4: Dynamic evolution of the AMGEN network, as time progresses the structure of the network evolves into a more complex pattern; Tier 1 structure changes – impacts upper tiers.

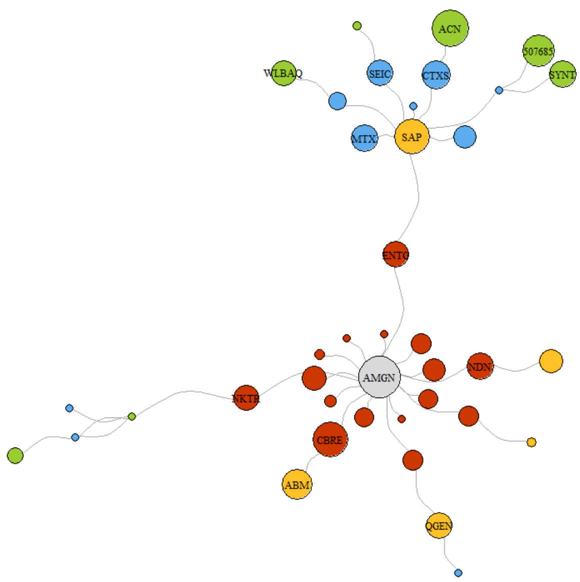


Figure 5: AMGEN 2005, Q1, companies in the supply network highlighted by tier and sized by revenue – top ten companies highlighted by name on the map.

Building layers in the network data

In addition to the general visualization of the node structure and the tiered relationships, we also wanted to understand the operation of the supply chain, hence the need for a geographic layer. In this visualization, we superimpose the nodes in the supply network (colored by tier) over a geographical map and use this map to explore node location and as an example relative size (revenue).

An example of the basic geographical layer of the network is shown in Figure 6. The blue node on the west coast is AMGEN, the focal node. Tier 1 suppliers are distributed mainly on the west coast with some presence on the east coast. The nodes are sized by revenue, for example, CBRE Group is one of the more significant suppliers in the network.

We can also provide additional layers of information which can be added to the basic dashboard shown in Figure 6. These include a structural dashboard (Figure 7) and a financial dashboard (Figure 8). For the financial dashboard, we provide three aspects of financial information including firm size measures, measures of financial performance as well as profitability. For the structural dashboard, we provide structural data such as the number of suppliers arranged by tier as well as node/edge characteristics.

By manipulating the information from the dashboards, we can then answer questions like, who are the key suppliers in the supply chain network and where are they located? What proportion of suppliers are located in a particular geographical region? If a disruption occurred, how would the supply network be affected? The structural and financial layers are still in the development phase and more enhancements are continually being added.

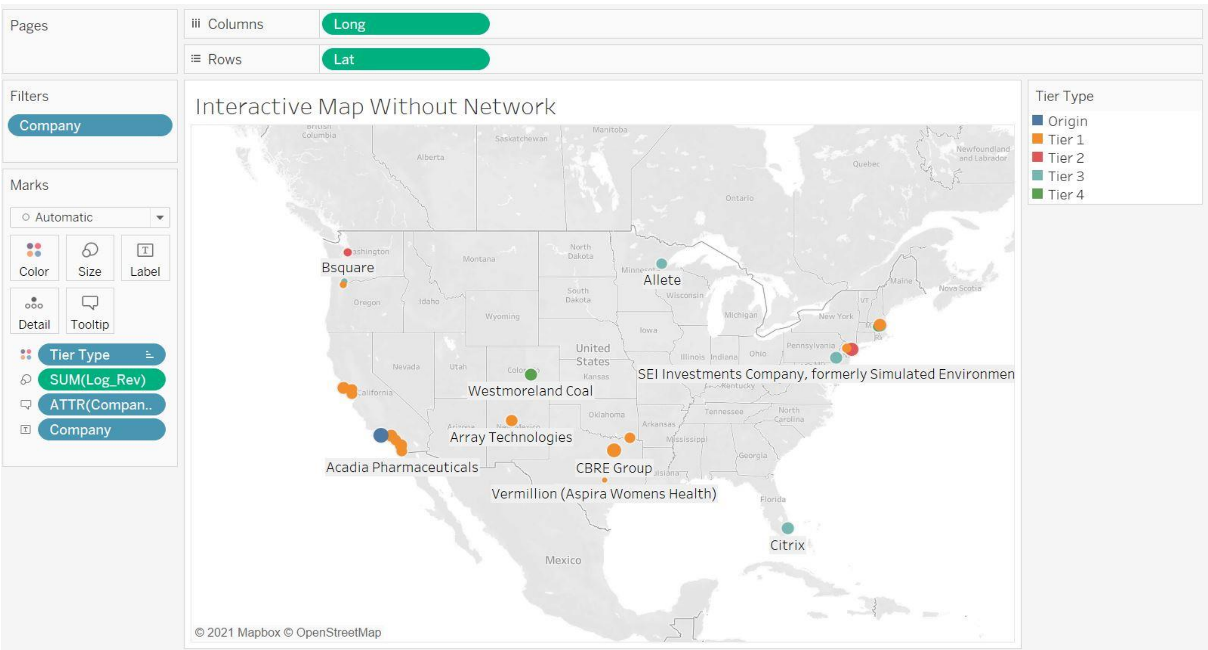


Figure 6: Geographical overview of the AMGN (blue node) network, 2005, Quarter 1.

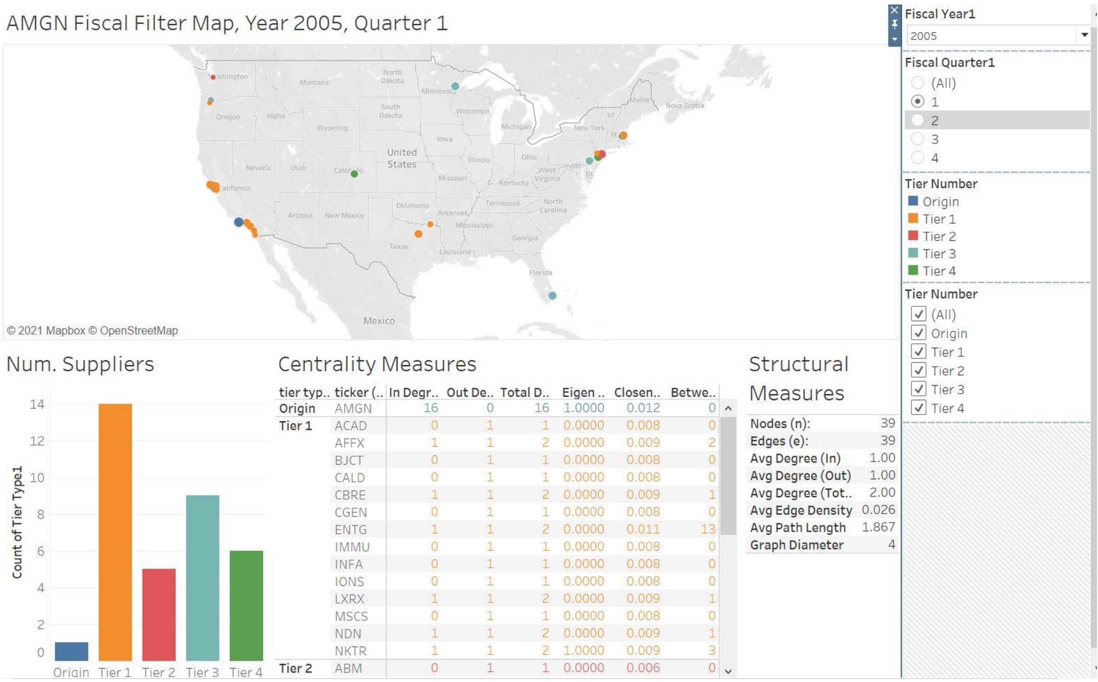


Figure 7: Structural dashboard of AMGN network, 2005, Q1

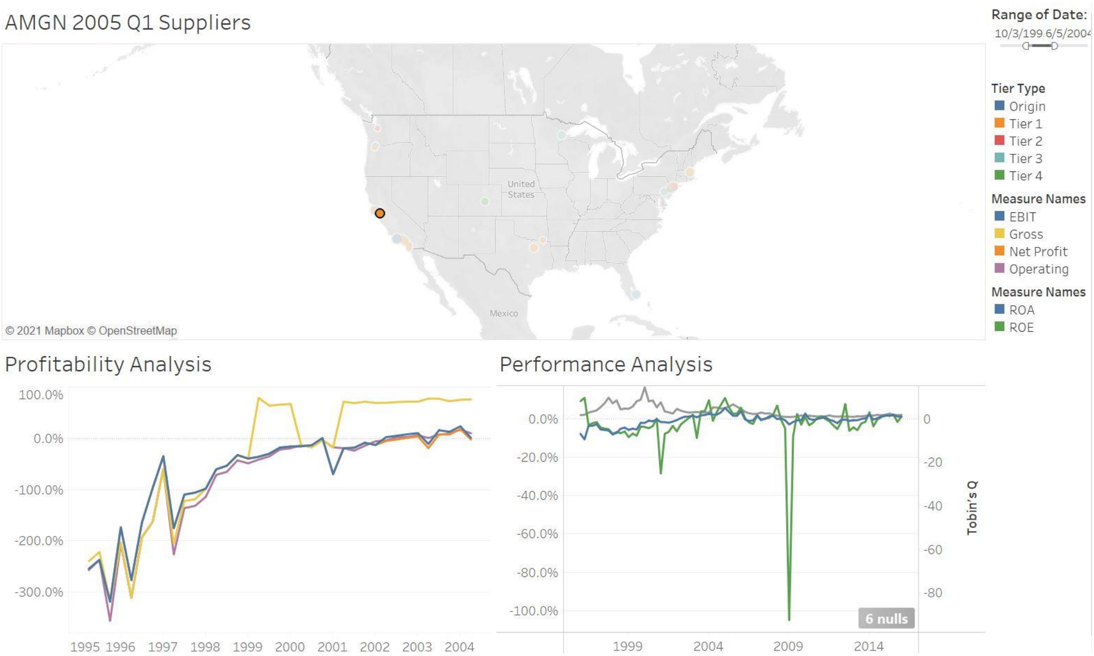


Figure 8: Financial dashboard of AMGN network, 2005, Q1

Network Structure Analytics

For the focal company's visualizations, we also calculated the corresponding structural metrics and recorded them by quarter and tier. We use several network metrics to characterize the topology of the supply network. These include the number of nodes, number of edges, average degree, network diameter, average path length and graph density. In the results section we describe a sample of data analyses that can be conducted using the empirical data. A sample construct of the quantitative data used in this research is shown in Table 1. The definition for each metric is provided in Appendix 1.

Year 2005	Node Count	Edge Count	Average Path Length	Average Degree	Network Diameter	Graph Density
Tier 1	16	16	1	1	1	0.059
Tier 2	6	6	1	1	1	0.045
Tier 3	10	10	1	1	1	0.064
Tier 4	7	8	1	1.14	1	0.051

Table 1: Structural Metrics for the AMGEN network, 2005, Quarter 1, Tier 1-4.

We describe the structural metrics and provide descriptive analytics for the AMGEN network. We evaluate the metrics over time to obtain an understanding of the evolution of the network. The structural metrics, when combined with the graphical visualizations demonstrated in Figures 3-5 yield information which explains the supply network operation. By combining financial, structural, and geographical information about the supply network, a true picture of the supply network can emerge. The information presented in the structural dashboard of Figure 7 is obtained from the analyses described in Figures 9-13.

In Figure 9, we view the relationship between node and edge counts as the network expands. In the first tier, these statistics correspond since typically the nodes are connected to the focal node in a concentric circle. But, as the network expands, the relationship between edge count to node count diverges. This is particularly noticeable from 2013, where the edge count exceeds the node count. The graphs show that there is a correspondence between the growth in AMGEN's revenue, highlighted in Figure 1, (2013 onwards), and network growth as shown in Figure 9. By monitoring the structural growth of a supply network, one can obtain an idea about the fiscal health of the focal network.

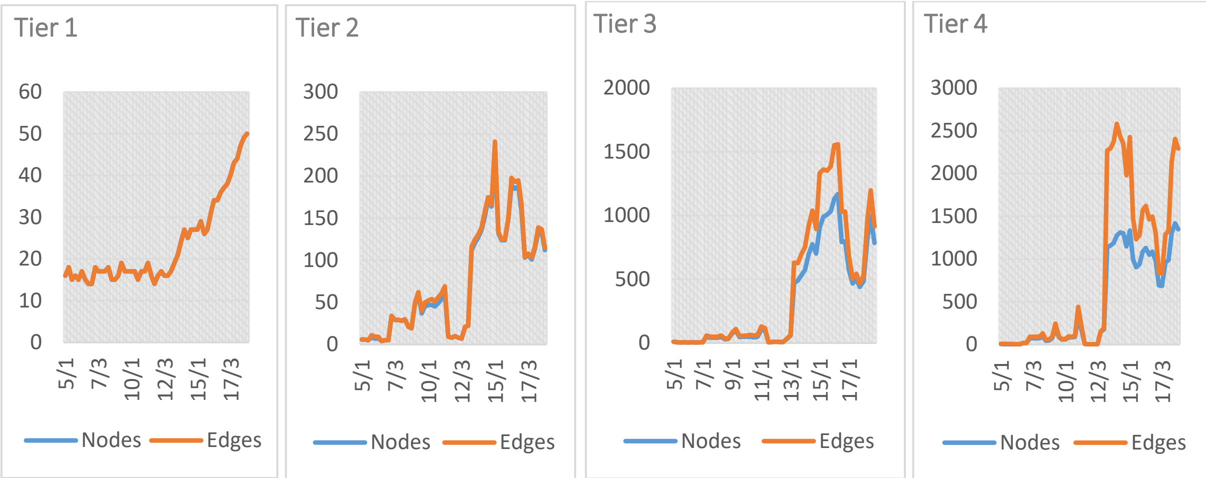


Figure 9: AMGEN Node and Edge count by Tier, 2005 – 2018; Source FACTSET.

The diameter of the network is the largest distance between any two firms in the network. As this number gets larger, it may be difficult to govern the supply network under a centralized regime. The diagrams in Figure 10 show the network expansion from this perspective; one can observe that with the addition of each tier, the network expands and the largest distance between any two firms in the network grows proportionately. In 2013, at tier 4, the diameter is at its largest at 17 hops but then scales back. This statistic is important since it has implications for network control in a centralized network. As the network diameter increases, it may be more difficult to control the network in its entirety. If the network diameter remains stable, (or within the same range of values), then network control may be achieved.

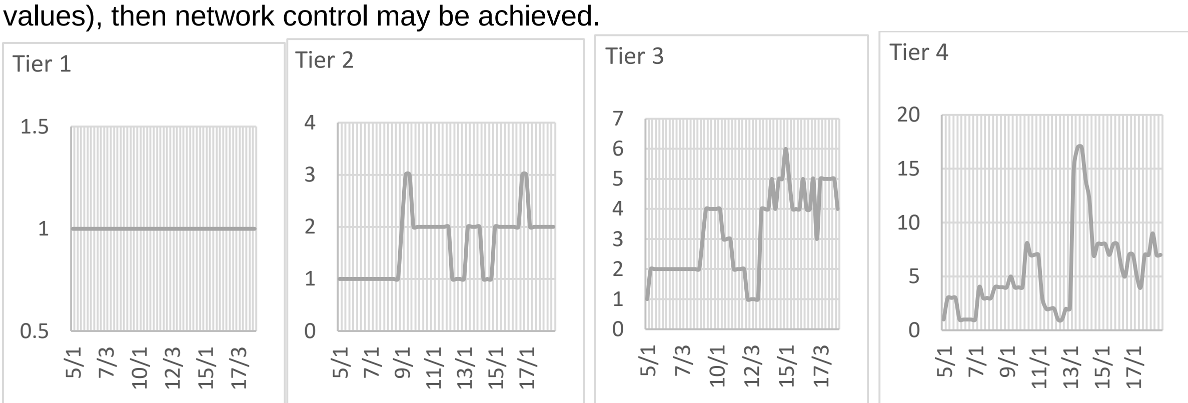


Figure 10: AMGEN Network Diameter by Tier 2005-2018

The average shortest path length provides a measure of how close the firms are to each other. In the diagrams in Figure 11, we can observe a similar trend to the network diameter. The average shortest path length increases as more tiers are included in the SCN. When the network is small (Tier 1 and Tier 2), this statistic remains stable hovering around the number 1 which implies that the firms in the network are closely located. As more tiers are included in the network, the average shortest path length grows commensurately indicating a more complex architecture among firms. If one compares Figure 5 with the data in Figure 11 one can deduce that the more complex structure in 2015 corresponds to an increase in the average shortest path length most noticeably as we examine the higher tiers.

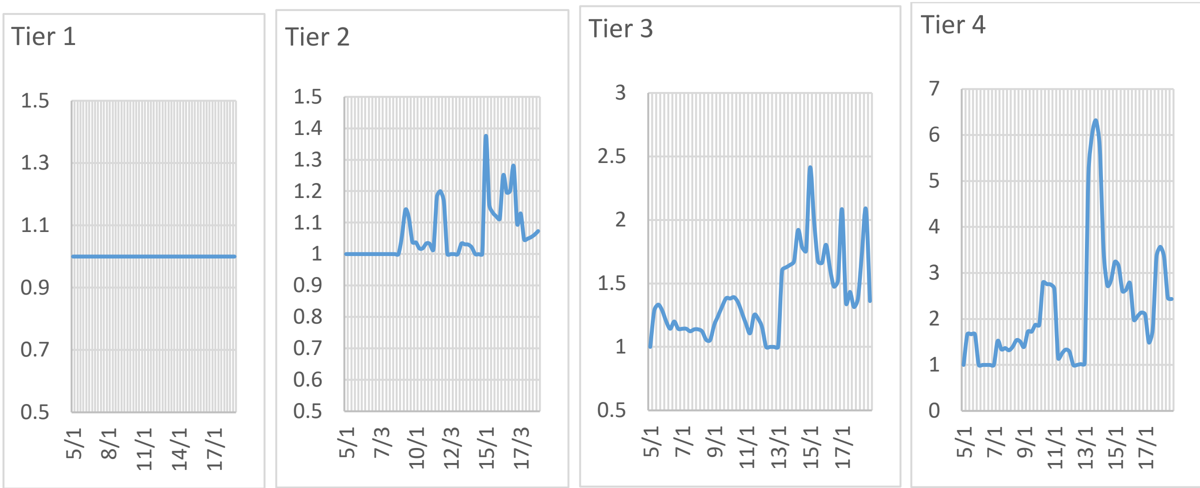


Figure 11: AMGEN Average Shortest Path Length by Tier 2005-2018

Since the average degree reflects the ratio of edge count to node count it follows that to some extent, average degree can also be used as an indicator of supply network expansion. When the average degree increases, this implies stronger inter-connectivity among the firms in the network. In fact, our previous research (Orenstein and Tang, 2021) indicated that in general, average degree is negatively correlated with firm performance, and that the overall performance of the Supply Chain Network (SCN) decreases as more connections form in the SCN. In Figure 12, we can confirm that, the average degree reaches a high in 2013 and then appears to fall. This decrease might be an indication of the sharp increases in financial performance highlighted in Figure 1, and the negative correlation we observed in our earlier research, (Orenstein and Tang, 2021).

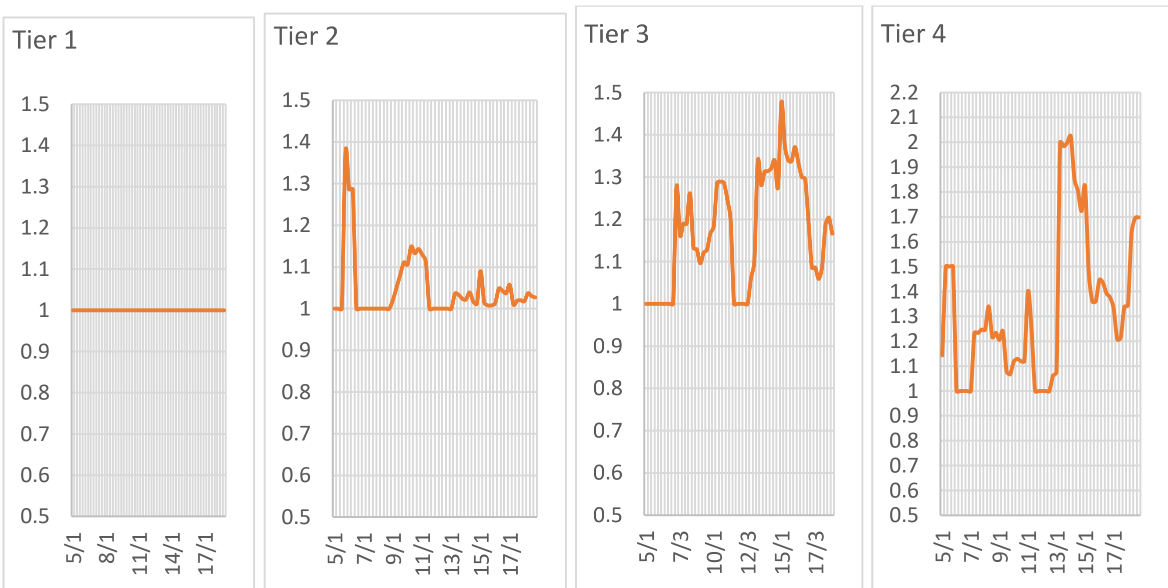


Figure 12: AMGEN Average Degree by Tier 2005-2018

Density of a SCN indicates the level of interconnectivity between the firms involved. SCNs with high density indicate good levels of connectivity between firms which can be favorable in terms of efficient information exchange and improved robustness due to redundancy and flexibility, (Sheffi and Rice, 2005). In Figure 13, the overall graph density increases as more tiers are included in the SCN, yet, unlike the other statistics of network diameter, average degree and average path length, graph density appears to be declining.

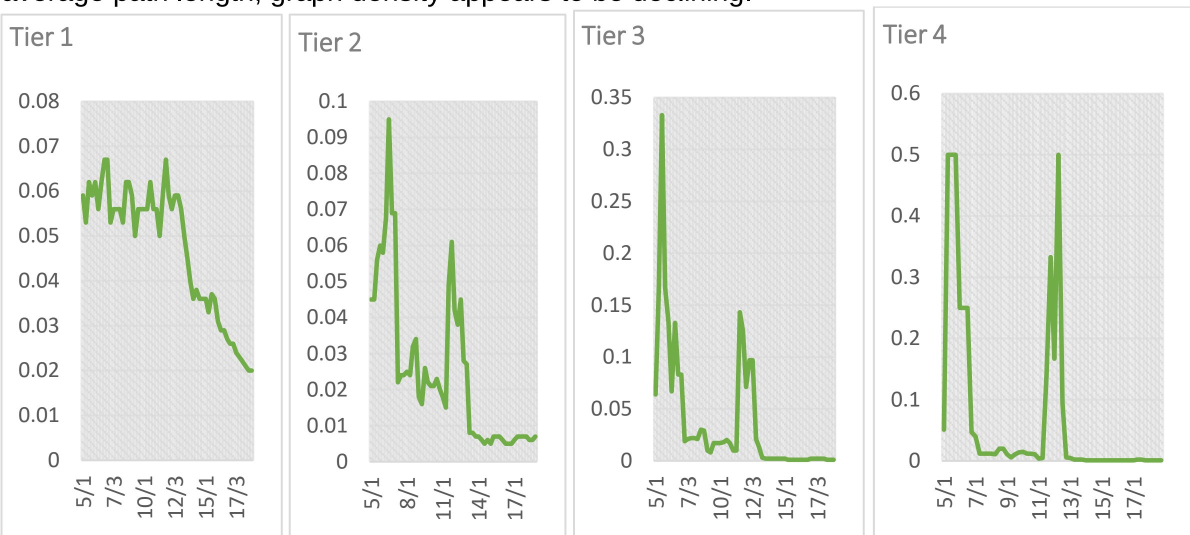


Figure 13: AMGEN Graph Density by Tier 2005-2018

CONCLUSIONS AND FURTHER WORK

In this paper, we described a framework which can be applied to the existing visual form of a supply network. The mechanism enables us to obtain some insight into the network operation. The idea was to introduce additional layers of information, including structural, geographic as well as financial. By superimposing these three dimensions on the existing network map, we are able to provide an understanding of the network that goes beyond the picture. Essentially, each dimension contributes a layer of knowledge to the visualization, thereby creating a measure of transparency in the original, somewhat opaque visualization. Using this three-pronged approach, the result is a comprehensive visualization of the supply chain network under consideration. In addition, the technique can easily be extended to study the dynamic behavior of the network.

In the paper, we used the AMGEN network to develop the overall mechanism which can then be applied to study the operation of the supply network. We started with a basic network map which highlighted the tiered relationships in the supply chain. The map was then enriched with geographical information which highlighted the locations of each of the firms in the network. Structural relationships were then calculated to provide a quantitative overview of the dynamic evolution of the networks. This information was then supplemented with financial data which helped identify the dominant nodes in the network. Both strands of information (financial and structural) are presented as part of the visualization resulting in a meaningful understanding of the supply chain network's operation. The framework that was developed in this paper is easily applicable to other supply network structures for which tiered relationship data is available.

Using the research tool developed in this paper, we would like to examine, to what extent, when there is growth and the network expands continuously by the addition of new nodes, do these new nodes attach preferentially to nodes that are already well connected? Is a large, well-known company more likely to attract new nodes or suppliers at a faster rate than a small, less-known company? Additionally, we plan to use the maps to explore the impact of COVID-19 on supply chains. For example, how does the COVID-19 pandemic affect the financial performance of companies with different supply chain networks. With a supply chain network framework in place, we can target these areas for future work.

REFERENCES

Zhu N., Zhang D., Wang W., Li X., Yang B., Song J., Huang B., Shi W., Lu R., Niu P., (2020) A novel coronavirus from patients with Pneumonia in China, 2019, N Engl. J Med 382(8): 727-733.

International Labor Organization (2019), Ending Child Labor, Forced Labor and Human Trafficking in Global Supply Chains, Alliance 8.7 Report. Available at https://www.ilo.org/ipec/Informationresources/WCMS_716930/lang--en/index.htm.

Cernansky, R. (2020), Where Fashion's Transparency Falls Short, Vogue Business, (retrieved April 30, 2020). Available at https://www.voguebusiness.com/sustainability/where-fashion-transparency-falls-short-raw-materials.

Norton T., Beier J., Shields L., Househam A., Bombis E., and Liew, D. (2014), A Guide To Traceability: A Practical Approach to Advance Sustainability in Global Supply Chains, https://www.bsr.org/reports/BSR_UNGC_Guide_to_Traceability.pdf;

Fair Labor Association (2017). Supply-chain Mapping, Transparency, Traceability – 1.0: A Practical Guide for Companies and Suppliers, Fair Labor Association (2017). Available at https://www.fairlabor.org/sites/default/files/traceability_guidance_paper.pdf.

Meiklejohn M., (2015) Supply Chain Transparency at Eileen Fisher; retrieved Feb 6th, 2015, https://www.youtube.com/watch?v=IIw5otZa8JQ&feature=youtu.be

Orenstein, P (2020), The changing landscape of supply chain networks: an empirical analysis of topological structure, INFOR: Information Systems and Operational Research, July 2020, pp 1-21.

Frost & Sullivan, (2019), Digitally Perfecting the Supply Chain, Whitepaper, downloaded from https://www.ibm.com/downloads/cas/ygbkmxjz

Vakil, B, (2020), Re-shoring starts with mapping suppliers, Inbound Logistics, Nov 2020.

Vakil B., (2021) Supply chain resiliency starts with supplier mapping, CSCMP Quarterly, Jan 2021.

Choi, T, Rodgers, D. and Vakil, B. (2020) "Coronavirus Is a Wake-Up Call for Supply Chain Management", https://hbr.org/2020/03/coronavirus-is-a-wake-up-call-for-supply-chain-management

S&P Global (2020), https://www.spglobal.com/en/research-insights/articles/hm-s-supply-chain-technology-investments-pay-off, retrieved Feb 4th 2020;

Orenstein P., and Tang H., (2022) An empirical analysis of the relationship between supply network structure and its financial performance, in preparation.

Orenstein P., (2020) The changing landscape of supply chain networks: an empirical analysis of topological structure, INFOR: Information Systems and Operational Research, DOI: 10.1080/03155986.2020.1785263;

Orenstein P., and Tang H., (2021) Identifying the relation between a supply chain network's structure and its overall financial performance, Operations and Supply Chain Management: An International Journal, to appear.

Sheffi, T. and Rice, J., (2005) A Supply Chain View of the Resilient Enterprise, MIT Sloan Management Review; Cambridge Vol. 47(1), pp 41-48.

APPENDIX A

These metric definitions are summarized below along with supply network context.

Mathematical Representation	**Contextual Definition**
Network Size *(N,L)*	The size of a given network is defined as the number of nodes and links and characterizes the overall scale of the network.
Degree $k_i = \sum_j a_{ij}$ where a_{ij} is any element of the adjacency matrix *A.*	Represents the number of connections a company has. For example, in a network, the company with the highest degree will likely be a key operator in the network.
Average Degree $\langle k \rangle = \frac{\sum_i k_i}{N}$ where *N* is total number of nodes in the network.	Indicative of how many connections a firm has. A high average degree implies strong inter-connectivity among the firms in the network.
Network diameter (*d*) $diameter = max_{i,j} l(i,j)$ where *l* is the number of hops traversed along the shortest path from node i to node j.	The diameter of the network is the largest distance between any two firms in the network. As this number gets larger, it may be difficult to govern the supply network under a centralized regime.
Average Path Length is defined as $m = \frac{1}{N(N-1)} \sum_{i \neq j} d(v_i, v_j)$ where $d(v_i, v_j)$ is the shortest path between nodes *i* and *j*.	The average shortest path length provides a measure of how close the firms are to each other.
Graph Density is defined as $D = \frac{\langle k \rangle}{N-1}$, where $\langle k \rangle$ is the mean degree of all nodes and N is the number of nodes in the network with $0 \leq D \leq 1$	Density of a SCN indicates the level of interconnectivity between the firms involved. SCNs with high density indicate good levels of connectivity between firms which can be favorable in terms of efficient information exchange and improved robustness due to redundancy and flexibility, (Sheffi and Rice, 2005).

Table A1: Node and network level metrics used to describe the supply network topology.

DECISION SCIENCES INSTITUTE

The Co-creation of Value in Online Business Education from a Student's Perspective

Abubaker Haddud
Department of Business Innovation and Strategy
Southwest Minnesota State University
Email: abu.haddud@smsu.edu

Anshuman Khare
Faculty of Business
Athabasca University
Email: anshuman@athabascau.ca

Terry Beckman
Faculty of Business
Athabasca University
Email: tbeckman@athabascau.ca

ABSTRACT

Universities around the world are embarking on online learning and delivery of courses. This instruction method has become a must have, rather than good to have for institutions. Online learning requires continuous improvement efforts to ensure online programs are delivered in an efficient manner. This research investigates the applicability of the service-dominant logic framework as a means to improve teaching and learning in an online graduate business education environment. While other research has looked at co-creation in education, this research goes further and examines four of the five axioms [excluding the first one] that are a part of the service dominant logic framework. An online survey consisting of questions related to the four chosen axioms was used to collect data from 143 online students enrolled in an executive online MBA program. The findings show that the service-domain logic can be used as an effective framework to better understand aspects related to the co-creation of value in online business education. The findings also confirmed key contributors to the concept of value co-creation such as learning environment, student-to-faculty interaction, and integration of available online resources. These findings show the value of applying the entire service dominant logic framework to improve the practice of education. This study is the first of its kind to utilize the complete service-dominant logic framework to better understand the concept of value co-creation and how to maximize the student's performance and outcomes in an online graduate business education.

KEYWORDS: Value co-creation, online graduate business education, service-dominant logic.

INTRODUCTION

The co-creation of value – a core component of the service-dominant (S-D) logic – has merit for education (e.g., Canallone et al. 2019; Jarvis et al. 2014; Kalafatis & Ledden 2013; Ng & Forbes 2009; Semeijn et al. 2011). Education is fundamentally a service exchange from which students acquire value. Students are involved in the creation of this value, that is, they co-create the value (e.g., Beckman & Khare, 2017; Chung & McLarney 2000), thus, one would expect the axioms of the S-D Logic to apply. Through a survey, we wish to take a closer look at S-D Logic framework presented by Vargo & Lusch (2004; 2008; 2016) and explore its value in online business education in a post-secondary context. In particular, four of the five axioms of service-dominant logic are examined in this context to determine the degree of applicability of the framework. This research intends to provide useful insights for the development of effective online business programs.

The service-dominant (S-D) logic, as introduced by Vargo and Lusch (2004, p. 9), is “a mindset, a lens through which to look at social and economic exchange phenomena so they can potentially be seen more clearly.” One of the core tenets of the S-D logic is the co-creation of value, where the customer is “a co-creator of value” (Vargo & Lusch, 2008, p. 7) and an enterprise/organization offers a value proposition; together, the customer and organization create value. However, this axiom is only one of the five that comprise the S-D Logic. The notion of co-creation has been empirically studied, and shown to be important in many contexts, including an educational context (e.g., Cavallone et al. 2019; Kalafatis & Ledden 2013; Ng & Forbes 2009). If the S-D Logic applies to an education context, the other axioms should also be relevant and important. The five axioms of the S-D logic from Vargo and Lusch (2016, p. 47) are:

- Service is the fundamental basis of exchange.
- Value is co-created by multiple actors, always including the beneficiary.
- All social and economic actors are resource integrators.
- Value is always uniquely and phenomenologically determined by the beneficiary.
- Value co-creation is coordinated through actor-generated institutions and institutional arrangements.

In this research we are primarily interested in axioms 2, 3, 4 and 5, all of which specifically deal with value. We assume that the first axiom holds in an educational context (Woodall, Hiller & Resnick 2014; Ng & Forbes 2009), thus we do not test for it. The fourth axiom seems to be looking at some of the sub-details around value creation, and thus is at a lower level of analysis than we are interested in for this research. It is, however, an interesting axiom, and we decided to address it in this research as well. Overall, we feel it is important to look at more than just value co-creation to understand if S-D Logic applies to education. Thus, this research seeks to answer the following three questions:

RQ1: What is the applicability of using the service-dominant logic framework in online business education context?

RQ2: What are the key factors contributing to the process of value creation in online business education?

RQ3: how can the value co-creation process in an online business education environment be maximized?

This paper is organized as follows: Section 2 provides a literature review about online business education and S-D Logic within this domain. One null and four alternative hypotheses are developed in this section. This section also provides the research framework diagram. Section 3 includes information about the research methodology and method used for data collection. Section 4 includes analyses of the collected data, results and findings, and discussions of these findings linking them to relevant literature. Finally, Section 5 discusses research conclusions, theoretical contribution and managerial implications, and research limitations and future research recommendations.

LITERATURE REVIEW

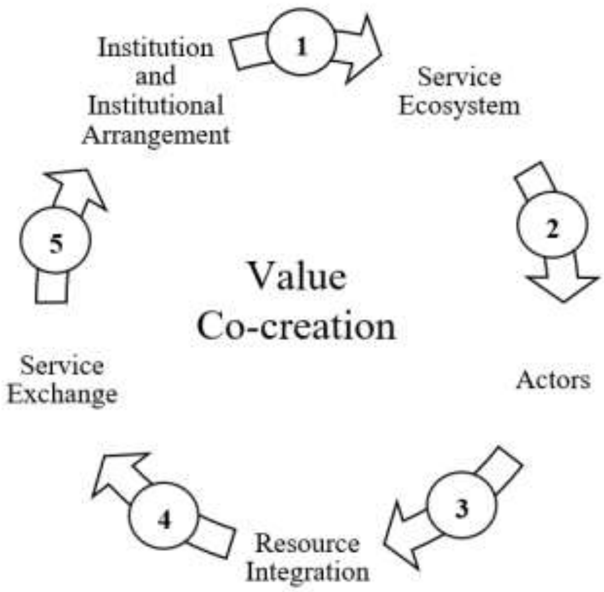
Online Business Education

Online education is defined as a form of distance education that uses computers and the Internet as the delivery mechanism, with at least 80% of the course content delivered online (Allen & Seaman, 2008; Shelton & Saltsman, 2005). The first undergraduate online education endeavor can be traced back to 1989 when the University of Phoenix began offering its online services (Kentnor, 2015). On the other hand, the first accredited online MBA program was offered by Aspen University in Denver, Colorado, United States back in 1987 (Hatfield and Spiller, 2017). In addition, the first online MBA for executives program was offered by Athabasca University in Canada in 1994. Online learning, in general, has evolved dramatically because of several factors including available technological tools, faster internet services, more people connected to the internet, more acceptance of online learning as a flexible method, more employers accepting online credentials, and more universities opting to offer more online programs. Almost since the first online MBA programs were launched researchers have been looking at how to make these programs effective (Arbaugh 2000). While much progress has occurred, there is always room for improvement. One framework that has potential for practical benefits in this context is the Service-Dominant Logic (Vargo & Lusch 2004).

Service-dominant logic and value co-creation

Service-Dominant Logic is a relatively recent perspective for understanding how value is created through service exchange in markets and economies (Lessard et al., 2019). This concept focuses on customer and value co-creation and provides a thorough concept of service system (Ghorsi et al., 2018). Service-Dominant logic is based on the assumption that no value can be obtained in isolation, but value is derived from exchanges, interactions, and collaborations (Gummesson et al., 2019). Service-Dominant logic is built upon five axioms, which are used to study a theoretical construct called service ecosystems (Vargo and Lusch 2016). The five axioms have been developed underlining that “value co-creation is coordinated through actor-generated institutions (i.e., norms, rules, symbols, meanings, etc.) and institutional arrangements” (Vargo and Lusch, 2016, p. 6). Figure 1 shows the five axioms of the S-D logic adopted from Vargo and Lusch (2017, p. 47):

Figure 1: The five axioms of the S-D logic

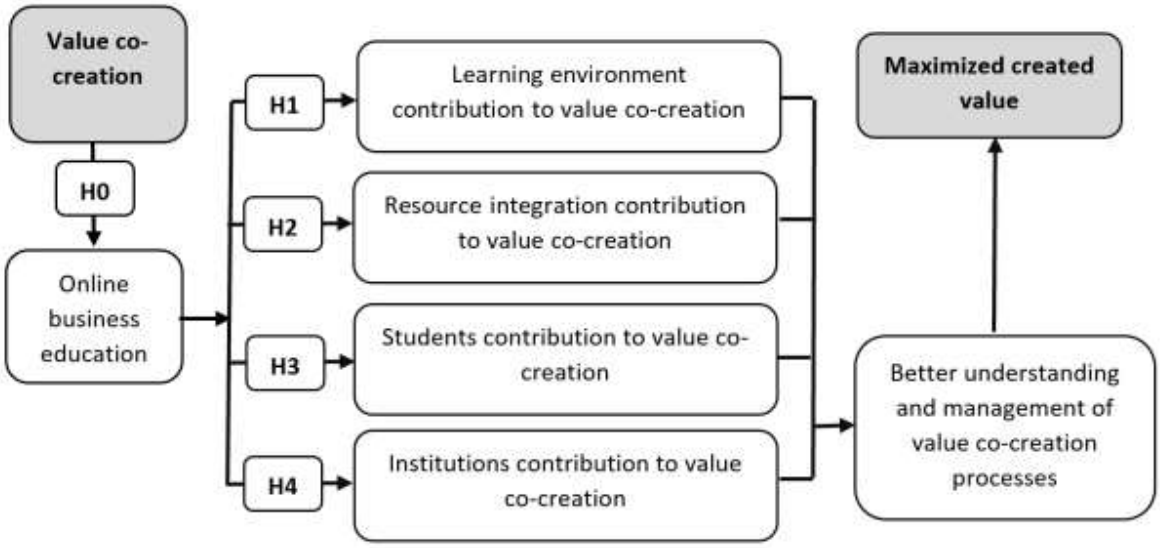


S-D Logic and Online Business Education

Assuming that education is a service context (Woodall, Hiller & Resnick 2014; Ng & Forbes, 2009), S-D Logic axioms should be applicable. Jarvis et al. (2014, p. 2) argue that "learners need to be active, participatory, and engaged in the learning process", and that this can be partly achieved through a flipped classroom, to which the S-D Logic framework is particularly appropriate. They maintain that the participation and engagement of the students in the flipped classroom is fundamental to the co-creation of value espoused by S-D Logic. While their (Jarvis et al. 2014) study focuses on the outcomes of engagement and the use of S-D Logic in a university, we look at the various aspects of the education experience of students and how those aspects fit into four of the S-D Logic axioms.

In this study, the first axiom was excluded, and survey questions were developed to address the remaining four. This research is based on the research framework shown in figure 2. One null and four alternative hypotheses have been developed to explore key aspects related to the concept of value co-creation in an online business education setting.

Figure 2: Research Framework



Research Hypotheses

H0: The framework of value co-creation presented in the service-dominant logic is not applicable for online business education.
H1: Learning environment contributes positively to the process of value co-creation.
H2: Resource integrations have direct impact on the quality of created value.
H3: Students play a key role in the process of value co-creation.
H4: institutions and institutional arrangements contribute to value co-creation processes.

RESEARCH METHODOLOGY

Instrument Development

An online survey was utilized to collect primary data from targeted student participants. The survey included the following parts: 1) General Questions (4 items); 2) Axiom 2 (5 constructs and a total of 30 items); 3) Axiom 3 (2 constructs and a total of 12 items); 3) Axiom 4 (2 constructs and a total of 19 items); and 4) Axiom 5 (1 construct and a total of 8 items). In parts 2, 3, 4, and 5, three main types of questions were used; 1) five-point level of influence scale; 2) four-point frequency scale; and 3) 1-5 rating scale.

Table 1: Survey development and structure (greyed parts included in this paper)

Survey parts	Theme of used constructs	Number of items	Used scale(s)
General Information		4	
Axiom 2: value co-creators in an online learning environment	Level of influence of students' interactions with different people at the university	11	Not at all influential (1); Slightly influential (2); Somewhat influential (3); Moderately influential (4); Extremely influential (5)
	Level of influence of completing different tasks on student's learning outcomes	5	Same as above
	Level of influence of different factors on student's performance in online study	6	Same as above
	Students actions during their online program	4	Never (1); Sometimes (2); Often (3); Very often (4)
	Students interaction with faculty during their online program	4	Never (1); Sometimes (2); Often (3); Very often (4)
Axiom 3: online learning social and economic resource integration	Level of influence of certain elements on students' learning experience	5	Not at all influential (1); Slightly influential (2); Somewhat influential (3); Moderately influential (4); Extremely influential (5)
	Some practices of online students	7	Never (1); Sometimes (2); Often (3); Very often (4)
Axiom 4: value in online education as determined by the student	Students' perception about course delivery	5	Never (1); Sometimes (2); Often (3); Very often (4)
	Reasons why students seek to obtain the qualification they are currently studying	8	1 (low); 2; 3; 4; 5 (high)
	Reasons why students chose to study online	6	1 (low); 2; 3; 4; 5 (high)
Axiom 5: institutions and institutional arrangements contribution to value co-creation	Level of institutional emphasis on certain common practices	8	Never (1); Sometimes (2); Often (3); Very often (4)

Data Collection, Sample, and Profile of Participants

Data was collected from current online executive MBA students in December 2018 using a web-based survey developed on Google Forms. Total number of received survey responses was 164. Out of this number, 21 were incomplete responses and they were excluded leaving a total of 143 usable survey responses. The profile of participants is provided in Table 2. The majority of participants were between 31 and 50 years of age (around 78%). Also, around 83% of participants had full-time jobs. Furthermore, around 57% of participants were males. Finally, around 80% of participants had already completed more than five courses in their current online programs. Therefore, the assumption is that most of the participated students have had adequate involvement in the online environment and are likely to be able to provide valid and credible answers to the survey questions.

Table 2: Profile of Participants

Question	Frequency (n=143)	Percentage (%)
Age Group		
• 20 - 30	7	4.90
• 31 - 40	48	33.57
• 41 - 50	62	43.36
• 51 - 60	24	16.78
• 61+	2	1.4
Employment Status		
• Full-time	119	83.22
• Full-time Homemaker	1	0.70
• Part-time	5	3.50
• Self Employed (hours vary)	7	4.90
• Unemployed	5	3.50
• Work from home full time	4	2.80
• Work from home part time	2	1.40
Gender		
• Male	81	56.64
• Female	62	43.36
Completed courses so far		
• 1	0	
• 2	6	4.2
• 3	12	8.4
• 4	10	7.0
• 5+	115	80.4

Data Analysis Methods

The IBM Statistical Package for the Social Sciences (SPSS) version 24 was used to analyze the data. First, the reliability and validity of the used survey were examined. Survey instrument consistency and reliability were examined. Cronbach alpha values were calculated for each construct as well. Also, an exploratory factor analysis was used to examine item and construct validity. Second, descriptive analyses were conducted to calculate frequency means for each item included in parts 2, 3, 4, and 5. Finally, Chi-square test (χ^2 test) was used for hypotheses testing.

Reliability and Validity of Research Instrument

Reliability is defined as the proportion of variance in observed test score that is related to true scores (Cronbach, 1951; McDonald, 2013) of the scale items in the same construct (DeVellis, 1991). Cronbach alpha values were used to examine the consistency and reliability of the items used in each of the survey constructs. As shown in Table 2, the Cronbach alpha coefficient values for the 10 constructs ranged from 0.813 to 0.892. Rivard and Huff (1988) suggest that Cronbach's values exceeding alpha coefficient of 0.7 threshold provide reliability evidence for internal consistency of the measurement scales. Therefore, it can be concluded that the survey instrument be considered as a reliable research tool that; 1) provides a thorough picture of the examined themes; 2) used items/statements that are directly related to the main theme of the construct; and 3) all of the used statements are inter-related.

Table 3: Used Survey Constructs' Validity and Reliability Tests Results

Used Scale	Number of constructs	Total No of items	Items factor loadings ranges	Construct Cronbach Alpha value
• **Axiom 2**: value co-creators in an online learning environment	5	30	0.542 to 0.838	0.843
• **Axiom 3:** online learning social and economic resource integration	2	12	0.468 to 0.845	0.837
• **Axiom 4:** value in online education as determined by the student	3	19	0.484 to 0.803	0.813
• **Axiom 5:** institutions and institutional arrangements contribution to value co-creation	1	8	0.626 to 0.845	0.892

Additionally, an exploratory factor analysis (Principal Component) was used to examine the construct validity of the used survey. This analysis helps to determine how, and to what extent, each item within survey constructs is linked to their underlined factors. Hair et al. (2006) state that the rule-of-thumb is that factor analysis values greater than 0.30 should be considered significant, values greater than 0.40 should be considered more important, and values that are 0.50 or greater should be considered very significant. As shown in Table 3, all the constructs had item factor loading values greater than 0.468. Therefore, the construct validity of the used survey construct is very significant.

Hypotheses Testing Method

Chi-square test (χ^2 test) was used to test the hypotheses in this study. The χ^2 test is widely used to examine data for differences, associations, and relationships to answer hypotheses (Waller, 2012). It compares observed frequencies (N) with the corresponding expected frequencies (N) under each of the five levels of agreement (Minaei-Bidgoli et al., 2004). The expected N can be determined by dividing the number of total usable responses (143 in this study) by the number of Likert-scale levels of agreement options (5 in this study). By dividing 143 by 5, the result is 28.6 and this was used as the expected N to calculate residual values for the survey constructs. For example, the first examined item at the top of the 'Overproduction' construct had actual responses (observed N) as follows: Not at all influential (5); Slightly influential (4); Somewhat influential (20); Moderately influential (52); Extremely influential (61). By subtracting the expected N value (28.6) from the observed N values, this results in residual values of -23.6, -24.6, -8.6, 23.4, and 32.4 and the highest residual value is under the fifth option 'Extremely influential'. Thus, this item was supported. Any examined item that had the highest residual value under 'Not at all influential'; 'Slightly influential'; or 'Somewhat influential' was considered as insignificant and the examined item was then be considered as 'Not Supported'.

RESULTS AND DISCUSSIONS

Axiom 2: Value Co-Creators in an Online Learning Environment

Level of influence of students' interactions with key people in the university on their learning experience and outcomes

Participants were asked to indicate the level of influence of 10 different people at the university on their learning experience and outcomes. As shown in Table 4, 'Students' came on the top of most influential people with a mean value of 4.13'. 'Course Faculty' came second with 3.65 mean value and 'Academic Advisors' came in the third place (mean = 3.31). The remaining seven groups had limited influence on students learning experience and outcome.

Table 4: Level of Influence of Students' Interactions with Different People at The University

People	Mean (n=143)	Cronbach α	Factor loading	Responses' frequency					Item is supported
				(1)	(2)	(3)	(4)	(5)	
1. Students	4.13	0.843	0.505	5	4	20	52	61	Yes
2. Academic advisors	3.31	0.837	0.355	17	26	24	43	32	Yes
3. Course Faculty	3.65	0.840	0.484	10	16	30	42	44	Yes
4. Other Faculty	1.96	0.824	0.412	70	21	32	11	5	No
5. Student services staff	2.37	0.819	0.621	46	35	28	18	12	No
6. Administrative staff and offices	2.59	0.813	0.673	31	38	40	18	15	No
7. Career services	1.49	0.834	0.646	100	20	13	6	1	No
8. Help desk	2.63	0.831	0.489	35	29	45	15	17	Yes
9. Registrar	2.31	0.815	0.651	44	41	29	18	10	No
10. Financial Aid Officer	1.34	0.842	0.588	111	13	11	2	2	No

Not at all influential (1); Slightly influential (2); Somewhat influential (3); Moderately influential (4); Extremely influential (5)

Level of influence of completing different tasks on student's learning outcomes

In this section, participants were asked to indicate the level of influence on five selected practices on their learning outcomes. As Table 5 shows, all of these practices either were reported to be either 'moderately influential' or 'extremely influential'.

Table 5: Level of Influence of Completing Different Tasks on Student's Learning Outcomes

		Mean (n=143)	Cronbach α	Factor loading	Responses' frequency (1)	(2)	(3)	(4)	(5)	Item is supported
1.	Doing individual assignments	4.62	0.561	0.788	0	2	4	40	97	Yes
2.	Work in groups	3.92	0.684	0.446	8	6	25	54	50	Yes
3.	Question professors and support staff	3.94	0.591	0.658	3	12	29	45	54	Yes
4.	Read, listen to or watch course materials	4.37	0.564	0.755	2	3	13	47	78	Yes
5.	Apply personal skills and knowledge	4.55	0.605	0.682	0	2	8	43	90	Yes

Not at all influential (1); Slightly influential (2); Somewhat influential (3); Moderately influential (4); Extremely influential (5)

Level of influence of different factors on students' performance in online study

In this section, participants were asked to indicate the level of influence of six previous factors on their performance in the current online study program. As indicated in Table 6, all the examined six factors were reported to be influential on students' performance. However, gained work experience from current and previous employment were reported as the top two factors.

Table 6: Level of Influence of Different Factors on Students' Performance in Online Study

		Mean (n=143)	Cronbach α	Factor loading	Responses' frequency (1)	(2)	(3)	(4)	(5)	Item is supported
1.	Previous academic qualifications	3.58	0.615	0.764	9	18	32	49	34	Yes
2.	Gained knowledge through previous training	3.86	0.506	0.806	2	9	38	54	39	Yes
3.	Gained work experience from previous employments	4.26	0.560	0.786	1	7	20	43	71	Yes
4.	Work experience from current job(s)	4.38	0.611	0.797	2	8	10	37	85	
5.	Other skills and competences e.g. English language, academic writing, etc.	4.07	0.583	0.717	4	6	28	41	62	Yes
6.	Support from family or/and friends	3.89	0.701	0.897	8	13	31	28	62	Yes

Not at all influential (1); Slightly influential (2); Somewhat influential (3); Moderately influential (4); Extremely influential (5)

Students' interactions with other students in the same online program

In this section, participants were asked to indicate the frequency of doing four identified practices during their current online program. As Table 7 shows, 'Prepared for exams by discussing or working through course material with other students' had the highest number of responses under the 'Never' answer option. 'Worked with other students on course projects or assignment' had the highest number of responses under 'Often' answer option. The other three examined practices had the highest number of responses under 'Sometimes'.

Table 7 Students Interactions with Other Students in The Same Online Program

		Mean (n=143)	Cronbach α	Factor loading	Responses' frequency				Item is supported
					(1)	(2)	(3)	(4)	
1.	Asked another student to help you understand course material	1.82	0.583	0.821	53	68	15	6	Yes
2.	Explained course material to one or more students	2.02	0.624	0.783	29	85	24	4	Yes
3.	Prepared for exams by discussing or working through course material with other students	1.43	0.685	0.675	96	35	7	4	Yes
4.	Worked with other students on course projects or assignments	2.80	0.701	0.656	5	51	53	33	Yes

Never (1); Sometimes (2); Often (3); Very often (4)

Students interaction with faculty during their online program

In this section, participants were asked to indicate the frequency of doing five practices during their studies. As Table 8 shows, all of these five examined practices had the highest number of responses under the 'Never' answer option.

Table 8 Students Interaction with Faculty During Their Online Program

		Mean (n=143)	Cronbach α	Factor loading	Responses' frequency				Item is supported
					(1)	(2)	(3)	(4)	
1.	Talked about career plans with a faculty member	1.23	0.509	0.682	117	21	3	2	No
2.	Worked with a faculty member on activities other than coursework (committees, student groups, etc.)	1.09	0.537	0.689	133	8	1	1	No
3.	Discussed course topics, ideas, or concepts with a	1.34	0.428	0.765	105	31	4	3	No

	faculty member outside of class								
4.	Discussed your academic performance with a faculty member	1.53	0.565	0.555	78	55	9	1	No

Never (1); Sometimes (2); Often (3); Very often (4)

As mentioned earlier, four hypotheses were developed for this research. The first one was *H1*: *Learning environment contributes positively to the process of value co-creation* and this section covered questions to help address this hypothesis. The results from this section were as follows:

1. The influence of ten people within the university has been examined and only four found to have a great impact on students' online learning experience. These were Students, Course Faculty, Academic Advisors, and Help desk.
2. In addition to this, students reported that their learning outcomes were impacted by; completing individual assignments; applying personal skills and knowledge; reading, listening or watching course materials; ask professors and support staff; and work in groups.
3. Furthermore, students reported that other external factors have positively contributed to enhancing their online performance that were related to current and previous work experience; other skills such as academic writing and English language; previous academic qualifications; and family support.
4. However, student-to-student interaction has been reported as limited. For example, limited experience was reported with discussing assignments with other students; explaining course materials to one or more students; or preparing for exams with other students. In contrast, students reported some student-to-student work on completing group assignments.
5. Finally, students reported very limited student-to-faculty interaction discussing aspects that are not purely course-related such as; talking about career plans; discussing course topics and concepts; or academic performance.

With the confirmed above three online environment aspects that influence students' learning outcomes and the limitations of two more; the first hypothesis "*Learning environment contributes positively to the process of value co-creation*" was <u>partially supported.</u>

Axiom 3: Online Learning Social and Economic Resource Integration

<u>Level of influence of certain elements on students' learning experience</u>

In this section, participants were asked to indicate the level of influence of five internal factors on their learning experience. As Table 9 shows, all of the examined five factors has the majority of responses under 'Moderately influential' and 'Extremely influential'.

Table 9 Level of Influence of Certain Elements on Students' Learning Experience

		Mean (n=143)	Cronbach α	Factor loading	Responses' frequency (1)	(2)	(3)	(4)	(5)	Item is supported
1.	Professors' shared course materials and their own experiences	3.91	0.727	0.612	7	8	29	45	54	Yes

2.	Shared students' knowledge and experiences	4.16	0.678	0.840	3	6	21	46	66	Yes
3.	Studying with students who come from different regions, cultures, and backgrounds	3.63	0.712	0.591	10	15	29	50	38	Yes
4.	Access to online resources (e.g. online library for reading materials, software needed to complete certain course tasks, etc.)	4.37	0.756	0.841	3	6	13	36	85	Yes
5.	Available tools to complete the required learning tasks and to communicate with professors, other students, and people in the institution e.g., Learning Management Systems	4.13	0.696	0.769	5	7	18	49	64	Yes

Not at all influential (1); Slightly influential (2); Somewhat influential (3); Moderately influential (4); Extremely influential (5)

<u>Some practices of online students</u>

In this section, participants were asked to share how often they do the seven examined practices. As Table 10 shows, more responses were under either 'Often' or 'Very often'. 'Connected ideas from your courses to your prior experiences and knowledge' examined item came in the first place and 'Combined ideas from different courses when completing assignments' came in the last place.

Table 10 Some Practices of Online Students

		Mean (n=143)	**Cronbach α**	**Factor loading**	**Responses' frequency**					**Item is supported**
					(1)	**(2)**	**(3)**	**(4)**		
1.	Combined ideas from different courses when completing assignments	3.15	0.858	0.674	1	29	59	54		Yes
2.	Connected your learning to societal problems or issues	3.34	0.834	0.822	0	13	68	61		Yes
3.	Included diverse perspectives (political, religious, racial/ethnic, gender, etc.) in course discussions or assignments	2.82	0.855	0.705	7	42	64	30		Yes
4.	Examined the strengths and weaknesses of your own views on a topic or issue	3.31	0.841	0.782	0	15	69	59		Yes

5.	Tried to better understand someone else's views by imagining how an issue looks from his or her perspective	3.29	0.839	0.793	0	17	68	58	Yes
6.	Learned something that changed the way you understand an issue or concept	3.35	0.844	0.761	0	16	61	66	Yes
7.	Connected ideas from your courses to your prior experiences and knowledge	3.57	0.854	0.710	0	6	49	88	Yes

Never (1); Sometimes (2); Often (3); Very often (4)

This section explored the influence of available resource integration on students online learning experience. The two above survey constructs included items that would help support or not support the second hypothesis that reads "*Resource integrations have direct impact on the quality of created value*". All the examined survey items under the above two constructs were rated as 'moderately influential' or 'extremely influential'. For example, the top rated most influential factor was students' online access to resources such as online library. Another influential factor was the shared professors' and students experience. Furthermore, students reported that connecting ideas from their courses to their prior experience and knowledge had a great influence on their learning experience that made them change the way they understand an issue or concept. Overall, and based on these findings, the second hypothesis "*Resource integrations have direct impact on the quality of created value*" was <u>fully supported</u>.

Axiom 4: Value in Online Education As Determined By The Student

<u>Students' perception about course delivery</u>

In this section. Participants were asked to report the frequency their faculty have done five practices. As Table 11 shows, more responses were under either 'Often' or 'Very often'. Faculty clearly explaining course goals and requirements came in the top place. Whereas 'used examples of illustrations to explain difficult points came in the last place.

Table 11 Students' Perception About Course Delivery

		Mean (n=143)	Cronbach α	Factor loading	Responses' frequency				Item is supported
					(1)	(2)	(3)	(4)	
1.	Clearly explained course goals and requirements	3.29	0.817	0.763	0	20	60	61	Yes
2.	Taught course sessions in an organized way	3.25	0.803	0.813	1	19	64	57	Yes
3.	Used examples or illustrations to explain difficult points	2.91	0.775	0.854	7	40	52	41	Yes
4.	Provided feedback on a draft or work in progress	2.74	0.813	0.788	23	33	43	42	Yes

5. Provided prompt and detailed feedback on tests or completed assignments	3.26	0.824	0.719	0	23	57	62		Yes

Never (1); Sometimes (2); Often (3); Very often (4)

Reasons for students seeking to obtain the qualification they are currently studying

In this section, participants were asked about preseasons why they seek to obtain the qualifications they are currently studying. As Table 12 shows, a 1 to 5 rating scale was used, and the top reason was ‘self-development’. This was followed by ‘satisfying personal ambitions’ and in the third place was ‘developing a specific skill set of competences. ‘Studying for the joy of learning’ and ‘Seeking better social status and esteem’ came in the last two places.

Table 12 Reasons Why Students Seek to Obtain the Qualification They Are Currently Studying

	Mean (n=143)	**Cronbach α**	**Factor loading**	**Responses' frequency**					**Item is supported**
				(1)	**(2)**	**(3)**	**(4)**	**(5)**	
1. Seeking better employment opportunities	4.29	0.735	0.712	8	4	17	22	91	Yes
2. Developing a specific skill set or competences	4.41	0.699	0.848	3	1	12	43	83	Yes
3. Increase knowledge of a particular disciplinary concept	4.24	0.695	0.826	3	1	22	47	69	Yes
4. Self-development	4.66	0.715	0.675	0	0	9	30	103	Yes
5. Satisfying personal ambitions	4.50	0.689	0.611	3	1	12	31	95	Yes
6. Studying for the joy of learning	3.61	0.715	0.528	14	16	27	38	46	Yes
7. Seeking better social status and esteem	2.84	0.673	0.702	35	32	17	33	23	Yes
8. Wanting more credentials	3.77	0.665	0.705	11	12	31	30	57	Yes

1 (low) --------- 5 (high)

Reason(s) for choosing to study online

In this section, participants were asked to report the reasons they chose to study online using a 1 to 5 rating scale. As Table 13 shows, ‘Flexible schedule to help balance work, family, and study commitments’ came in the first place. This was followed by ‘Able to attend class anytime’ and ‘Ability to attend class anywhere’. Whereas ‘Financial aspects e.g. less overall program cost compared to face-to-face offerings, funding, discounts, etc.’ came in the last place.

Table 13: Reasons Why Students Chose to Study Online

	Mean (n=143)	Cronbach α	Factor loading	Responses' frequency (1)	(2)	(3)	(4)	(5)	Item is supported
1. Interesting method of learning	2.79	0.502	0.592	34	23	43	21	21	Yes
2. Flexible schedule to help balance work, family, and study commitments	4.89	0.594	0.522	0	0	2	12	128	Yes
3. Financial aspects e.g. less overall program cost compared to face-to-face offerings, funding, discounts, etc.	2.98	0.61	0.404	35	20	33	19	34	Yes
4. Able to "attend" class anytime	4.62	0.531	0.724	4	2	7	17	112	
5. Ability to "attend" class from anywhere.	4.62	0.564	0.680	5	2	5	18	112	
6. Opportunity to interact with dispersed and heterogeneous student group	3.35	0.506	0.648	23	17	32	26	44	

1 (low) --------- 5 (high)

This section was dedicated to exploring the concept of value in online education as determined by students. Students reported that their professors did a good job in the delivery and instruction of courses by clearly explaining course goals and requirements; taught course sessions in an organized way; and provided prompt and detailed feedback on tests and completed assignments. Furthermore, the students saw great value in taking online education programs as self-development endeavors; increase knowledge of a particular disciplinary concept; satisfying personal ambitions; developing a specific skill set of competences; and seeking better employment opportunities. In addition, students reported that several key reasons were behind choosing to study online: flexible schedule to help balance work, family, and study commitments; ability to attend classes anytime and anywhere; and the opportunity to interact with dispersed and heterogeneous student group. Therefore, students saw a great value in online education and the third hypothesis "*Students play a key role on the process of value co-creation*" was <u>*fully supported*</u>.

Axiom 5: Institutions and Institutional Arrangements Contribution to Value Co-Creation

<u>Level of institutional emphasis on certain common practices</u>

In this section, participants were asked to report the level of institutional emphasis on eight common practices. As Table 14 shows, the first four practices had mean value more than 2. Whereas the rest had mean values below 2.

Table 14: Level of Institutional Emphasis on Certain Common Practices

	Mean (n=143)	Cronbach α	Factor loading	Responses' frequency				Item is supported
				(1)	(2)	(3)	(4)	
1. Providing support to help students succeed academically	2.80	0.881	0.739	3	55	51	33	Yes
2. Using learning support services (tutoring services, writing center, etc.)	2.31	0.883	0.685	28	59	38	17	Yes
3. Encouraging contact among students from different backgrounds (social, racial/ethnic, religious, etc.)	2.53	0.892	0.708	23	23	23	23	Yes
4. Providing opportunities to be involved socially	2.17	0.876	0.674	23	57	25	36	Yes
5. Providing support for your overall well-being (recreation, health care, counseling, etc.)	1.92	0.873	0.728	55	54	24	9	No
6. Helping you manage your non-academic responsibilities (work, family, etc.)	1.71	0.875	0.626	79	36	16	10	No
7. Attending campus activities and events (performing arts, athletic events, etc.)	1.63	0.877	0.777	74	52	9	6	No
8. Attending events that address important social, economic, or political issues	1.68	0.874	0.845	69	55	11	6	No

Never (1); Sometimes (2); Often (3); Very often (4)

This section was dedicated to exploring institutions and institutional arrangements contribute to value co-creation processes in an online learning environment. Out of eight explored aspects; four items were supported and four were not. Providing support to students to succeed academically and encouraging them to contact other students with different background came on the top of the supported list. Because the supported four items had relatively low rating and the other four were not supported; we believe the fourth hypothesis that reads "*institutions and institutional arrangements contribute to value co-creation processes*" <u>was not supported</u>.

CONCLUSIONS

Conclusion

This research sought to explore the applicability of the S-D Logic in an online business graduate learning environment. Four main hypotheses were developed and tested. As it is indicated in

the previous section; the first hypothesis "*Learning environment contributes positively to the process of value co-creation*" was partially supported. The second hypothesis "*Resource integrations have direct impact on the quality of created value*" was fully supported. The third hypothesis "*Students play a key role on the process of value co-creation*" was fully supported. Finally, the fourth hypothesis, "*institutions and institutional arrangements contribute to value co-creation processes*" *was not supported*.
The findings from this research are very rich and can be further presented in answering the three main research questions developed for this research. The first research question reads "*What is the applicability of using the service-dominant logic framework in online business education context?*". Our findings indicate that the S-D logic is a suitable framework that can be used to better understand factors impacting students' online learning experience. The framework helps institutions to better identify factors and areas that clearly contribute to the value co-creation and those areas that need more attention to maximize students learning outcomes.
The second research question reads "*What are the key factors contributing to the process of value creation in online business education?*" Findings from this research indicated that learning environments, students' interactions, and available online resource interactions as key contributors to the process of creating value for student in online business education.
The third research question reads "*how can the value co-creation process in an online business education environment be maximized*?" Based on the findings from this research, we decided to highlight any examined survey item that had an overall mean value less than 3.0 as potential areas for improvement to maximize students' value from their online studies. Examples of these areas include: 1) more interactions with the online students by student services; administration staff; help desk; registrar; career services; and financial aid. 2) Encouraging more student's interaction with other students such as creating shared online spaces where they can discuss course materials, prepare for exams, and help each other share information and understand what the expectations and requirements are. 3) Foster more student-to-faculty interaction discussing aspects such as career plans; academic performance; course topics and ideas; and activities such as committees and student groups. 4) Faculty providing examples or illustrations to explain difficult points and perhaps providing feedback on draft or work-in-progress.

Theoretical and Practical Contributions

The theoretical contributions of this research are three-fold: first, providing insightful information about the suitability of using the S-D logic framework to better understand graduate business online programs. Second, the finding from this research provides insights about the influence of key parts of students' performance and learning outcomes and the concept of value co-creation in an online business learning environment. Third, the study provides valuable information about maximizing the process of value co-creation through the identification of aspects that need more attention to better serve the online students.

This research has clear practical contributions that can be adopted by universities offering graduate online business programs. For example, universities can use the findings from this study to better understand how to efficiently manage the delivery online business programs particularly at the graduate level. The findings can be used as a framework to identify the required resources and success factors that will positively contribute to the success of their online programs. The adoption of such a framework provides a visual of the entire online operation and the key processes to help universities conduct thorough assessments of the performance of each of the involved players. The findings of this research can greatly help universities identify where the under-performing factors need to develop strategies to overcome any challenges. Such visualization and identification of what needs to be done should help universities better serve their students and maximize the value they get out of their online programs.

Limitations and Suggested Future Research

This research was limited to one online graduate business program at one university. We would suggest conducting a similar research on: different learning programs; multiple universities; and undergraduate programs as well. This would help verify the findings of this research and to see if these can be generalized to wider online learning communities. It is also suggested to conduct similar research exploring value co-creation aspects from the staff and administrators' point of view and compare the results with this research's results. Doing so will help visualizing how the value creation process is determined and perceived by all the involved key players. Finally, it may be worth conducting similar research with programs that use different virtual learning systems to identify variance due to learning systems; trying to determine if the used platforms have different influence on the overall students' online learning experience and on the concept of value co-creation.

REFERENCES

Allen, E. I., & Seaman, J. (2008). *Staying the Course - Online Education in the United States*. Sloan Consortium. https://eric.ed.gov/?id=ED529698

Arbaugh, J. B. (2000). Virtual Classroom Characteristics and Student Satisfaction with Internet-Based MBA Courses. *Journal of Management Education*, *24*(1), 32–54. https://doi.org/10.1177/105256290002400104

Beckman, T., & Khare, A. (2018). A Service-Dominant Logic and Value Co-creation Approach for Online Business Education. In A. Khare & D. Hurst (Eds.), *On the Line* (pp. 21–35). Springer.

Cavallone, M., Ciasullo, M. V., Douglas, J., & Palumbo, R. (2019). Framing higher education quality from a business perspective: setting the conditions for value co-creation. *Studies in Higher Education*, 1–13. https://doi.org/10.1080/03075079.2019.1672644

Chung, E., & McLarney, C. (2000). The Classroom as a Service Encounter: Suggestions for Value Creation. *Journal of Management Education*, *24*(4), 484–500. https://doi.org/10.1177/105256290002400407

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555
DeVellis, R. F. (2016). Scale Development Theory and Applications (Fourth Edition Robert). *SAGE Publication*.

Ghorsi, H., Siddiqui, A., Basharat, T., Anwer, S., & Ahmad, H. (2018). Service System of Easypaisa and MobiCash: Funds Transfer System via Telecommunication Companies in Context of Service Dominant Logic. *Journal of Service Science Research*, *10*(2), 249–273. https://doi.org/10.1007/s12927-018-0009-7

Gummesson, E., Sarno, D., Carrubbo, L., & Sirianni, C. (2019). Contributing to Sustainable Healthcare Systems with Case Theory. *International Journal of Business and Management*, *14*(2), 34–47. https://doi.org/10.5539/ijbm.v14n2p34

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Pearson Prentice Hall.

Hatfield, R., & Spiller, S. (2017). MBA Program at the crossroads: considering program innovations for closer customer-student engagement. *Global Journal of Business Pedagogy*, *1*(1), 28–39.

Jarvis, W., & Halvorson, W. (2014). A Large Class Engagement (LCE) Model Based on Service-Dominant Logic (SDL) and Flipped Classrooms. *Education Research and Perspectives*, *41*(1), 1–24.

Kalafatis, S., & Ledden, L. (2013). Carry-over effects in perceptions of educational value. *Studies in Higher Education*, *38*(10), 1540–1561. https://doi.org/10.1080/03075079.2011.643862

Kentnor, H. (2015). *Digital Commons @ DU Sturm College of Law : Faculty Scholarship Distance Education and the Evolution of Online Learning in the United States*. *17*(1), 22–34.

Lessard, L., Amyot, D., Aswad, O., & Mouttham, A. (2019). Expanding the nature and scope of requirements for service systems through Service-Dominant Logic: the case of a telemonitoring service. *Requirements Engineering*, *0123456789*. https://doi.org/10.1007/s00766-019-00322-z

McDonald, R. P. (2013). *Test Theory: A unified treatment*. Psychology Press.

Minaei-Bidgoli, B., Tan, P. N., & Punch, W. F. (2004). Mining interesting contrast rules for a web-based educational system. *Proceedings of the 2004 International Conference on Machine Learning and Applications, ICMLA '04*. https://doi.org/10.1109/icmla.2004.1383530

Ng, I. C. L., & Forbes, J. (2009). Education as service: The understanding of university experience through the service logic. *Journal of Marketing for Higher Education*, *19*(1), 38–64. https://doi.org/10.1080/08841240902904703

Rivard, S., & Huff, S. L. (1988). Factors of Success for End-User Computing. *Communications of the ACM*, *31*(5), 552–561. https://doi.org/10.1145/42411.42418

Semeijn, J., Semeijn, J. H., & Caniëls, M. C. J. (2011). How Should Service-Dominated Logic Be Applied to Business Education? In P. Van den Bossche, W. H. Gijselaers, & R. G. Milter (Eds.), *Building Learning Experiences in a Changing World* (pp. 45–54). Springer. https://doi.org/10.1007/978-94-007-0802-0_4

Shelton, K., & Saltsman, G. (2005). *An Administrator's Guide to Online Education*. IAP.

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, *68*(1), 1–17. https://doi.org/10.1509/jmkg.68.1.1.24036

Vargo, S. L., & Lusch, R. F. (2008). Service-dominant logic: Continuing the evolution. *Journal of the Academy of Marketing Science*, *36*(1), 1–10. https://doi.org/10.1007/s11747-007-0069-6

Vargo, S. L., & Lusch, R. F. (2016). Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science*, *44*(1), 5–23. https://doi.org/10.1007/s11747-015-0456-3

Woodall, T., Hiller, A., & Resnick, S. (2014). Making sense of higher education: Students as consumers and the value of the university experience. *Studies in Higher Education*, *39*(1), 48–67. https://doi.org/10.1080/03075079.2011.648373

DECISION SCIENCES INSTITUTE

Governance to Support Big Data Sharing in Supply Chain Management: A Design Science Approach

Wael Elsheikh
Maastricht University
Email: wael.elsheikh@insead.edu

Sherwat E. Ibrahim
American University in Cairo
Email: sherwat@aucegypt.edu

Ahmed Elragal
Luleå University of Technology
Email: ahmed.elragal@ltu.se

ABSTRACT

Retail businesses are now facing significant and obvious shifts and evolution in technology and innovation related to IoT and Industry 4.0. A move to more digitalized forms of ordering and processing is expected. However, even with the obvious trends and benefits of information sharing, retailers keep guarding their data and findings, which puts a challenge on retailer-manufacturer collaboration initiatives especially when it comes to Big Data Analytics (BDA) purposes. This paper raises issues related to the governance of information sharing initiatives related to BDA in retailer SCs, and proposes socio-technical framework on how Big Data Sharing (BDS) can be implemented.

KEYWORDS: Retail Supply Chains, Big Data, Data Analytics, Information Sharing, Inter-organizational Relational Governance, Design Science Research

INTRODUCTION

We live in the data era. Organizations are surrounded by sheer amounts of data, which originated from internal sources such as ERP, CRM, or SCM systems, as well as external sources such as social media and IoT systems e.g., weather and humidity sensors, RFID tags, etc. The number of interconnected devices is estimated to reach the 24 billion by end of 2020, which leads to exponential growth of data (Dev, et al., 2018). Increasing number of interconnected objects enable a new form of information transparency and paves the way for new applications of these technologies in supply chain and other sectors (Hermann, et al., 2016) (Dev, et al., 2018). The accelerating change in the ICT field drives the need for a change in organizations' behavior and its SCM (Huan, et al., 2004).
The rise of social media, smart mobile devices, IoT, IoE and Industry 4.0 has led to the growth in quantity and diversity of the amount of data generated and stored worldwide (Waller & Fawcett, 2013) (Feki, et al., 2016). Around 80% of this data is unstructured and is hard to be processed and analyzed with the conventional and traditional methods and tools (Waller & Fawcett, 2013) (Feki, et al., 2016). BDA is a powerful tool that can handle such data, but it cannot be separated from the relevant domain knowledge, in other words, to build a model both analytical methodologies and business-related knowledge should be used appropriately together (Waller & Fawcett, 2013). BDA could be performed either in real-time mode where there is a constant change in data and/or information, or offline mode where the results are needed in a very short time (Addo-Tenkorang & Helo, 2016).

IoT and Industry 4.0 are expected to affect how business is operated, managed, and controlled. The business environment is facing significant and obvious shifts and evolution in technology and innovation, which stimulate and suggest changing the SC of different organizations. Although, the magnitude of the effect may vary between the different business sectors and markets, but it should not be neglected.

These revolutionary changes enforce the need to adopt BDA in SC, which is expected to increase the organization's ability to compete in and benefit of the currently witnessed technological evolution, by aggregating raw data from different sources into higher-value context information and interpretation using Data Analytics (Hermann, et al., 2016). Advancement in Data Analytics methods increase the efficiency and the effectiveness of the operational performance and decision-making. This what defines the basis as well as the trigger of this research work, focusing on proper analysis, design, and implementation.

LITERATURE REVIEW

The governance of BD is crucial to the success of any organization that deals with huge amounts of data (Al-Badi, et al., 2018). BDG refers to processes, policies and procedures, adopted in order to manage the data asset, which commonly accomplished through a combination of people and process, with technology used to simplify and automate aspects of the process (Shah, 2017). BDG defines how to apply policies and regulations, and reflect proper data roles, such as data steward and data ownership, through the data lifecycle (Blazquez & Domenech, 2018) (Firican, 2017). BDG is not just about technology, it further includes those who are responsible for data asset along with the technology (XenonStack, 2018). Intra-organizational BDG enable sharing data while complying with current regulations, security, and privacy policies, and creates (Value-Adding) by providing the right data to the right person at the right time focusing on coordinating the efforts toward achieving common objectives (Addo-Tenkorang & Helo, 2016) (Al-Badi, et al., 2018) (Abraham, et al., 2019). On the other hand, Inter-organizational scope determines BDG between two or more entities; such as: suppliers, buyers, industry or cross industry peers, etc. focusing on coordinating the efforts toward achieving common objectives (Abraham, et al., 2019). To overcome issues associated with BDG across different entities, companies need to rely on BDG frameworks to identify related Governance Mechanisms; such as: Data Integration and Usage Policies, Data Exchange Standards, Processes for Interaction and Collaboration, Data Sharing Agreements, etc. (Abraham, et al., 2019).

Big Data Sharing

Information and data sharing between the different SC entities is not new to SCM, and it has been proven as an effective tool to increase the overall performance of the SC and the much-needed visibility in overcoming the Bullwhip effect (Lummus & Vokurka, 1999) (Grean & Shaw, 2002) (Tan, et al., 2002) (Yılmaz, et al., 2016). It enhances and affects the ability of collaborative partners to execute business transactions in a way that accomplishes the obligations of shared business processes in a competent manner (Cao & Zhang, 2011) (Liao, et al., 2017). Suggested change to adopt BDA in SC – when focusing on SC collaboration and integration – is expected to increase the organization's ability to compete in and benefit of the currently witnessed technological evolution. Therefore, it requires planning, rules, practices as well as standard operating procedures that guide the collaborative relationship (Zhu, et al., 2017).

BD sources across the supply chain can be classified based on the supply chain entities where the data were generated, into: upstream (i.e. suppliers), intermediate stream (i.e. manufacturers), consolidation points (i.e. warehousing), and downstream (i.e. logistics, distribution and/or retail) (Addo-Tenkorang & Helo, 2016). (Rozados & Tjahjono, 2014) in a comprehensive literature review identified more than 50 Supply Chain BD related sources. They classified these data according to Volume, Velocity, and Variety – the three Big Data levers. Most of the identified BD sources were found to be used only in the areas where they were first generated, creating additionally unfavorable data silos and barriers for implementation of BDA across the SC.

The SCM literature recognizes that achieving BD Sharing is challenging. Assembling different data sources and making the data accessible to different supply chain entities is a concern for the BDA implementation (Kaipia, et al., 2017). Most of the BD acquired from the different sources of the SC remain to be unstructured which makes it difficult to analyze using the traditional analytical tools (Lee, 2018). Yet, benefits are expected to be significant in terms of visibility and the ability to make informed timely decisions resulting from the access of huge data volumes, from different sources, in different formats, in a shorter time (Addo-Tenkorang & Helo, 2016). Ease of access to better data could facilitate improved understanding of supply chain operational problems. BDA is expected to provide an enabling value-adding platform that can affect operational efficiencies, enhance customer experience, and improve the process of new product development (Addo-Tenkorang & Helo, 2016). Identifying opportunities for BD sharing is considered a first step in the development of this uprising and a key milestone to reaping these benefits.

On the other hand, enablers and obstacles of the transformational process of BDA and BD sharing adoption and implementation remains unclear (Nguyen, et al., 2017). Successful implementation should be ensured by understanding the impact of adopting the new business model on the entire supply chain performance, and by defining the required contribution from different stakeholders and how they will be affected (Nguyen, et al., 2017). Therefore, influence of such Digital Transformation should discuss more in depth the evolution of the key pillars; such as: data sharing, decision synchronization, ownership of the collaborative process, and incentive alignment, as it will affect the way that the SC governance should be managed, and influence horizontal and vertical collaborative relationships.

Big Data Sharing and Governance

Inter-organizational relationships are challenged with the increased risk of losing control over data, lower data security and privacy, and lower data reliability, among many other obstacles (Wang & Wei, 2007) (Abraham, et al., 2019). Combinations of contractual and relational aspects are needed to make the governance structure effective (Mahapatra et al. 2010). Despite of the significant attention received by inter-organizational governance theories and mechanisms in SCM research, very few articles in literature have considered the impact of data sharing on SC governance. SCM Data and Information Sharing collaboration constructs like power balance, perceived value, decision synchronization, and incentive alignment are all necessary to create a situation of “Digital Trust” that would allow for the expansion of BD sharing and collaboration.

The choice for governance mechanisms is largely influenced by contextual factors (Bonatto, et al., 2020). It is important to identify the differences in context that may affect expected benefits, and to assess the attractiveness of BD application possibilities among stakeholders. Outlining best practices to take advantage of BD in real-life operations’ context can be useful to efficiently accomplish BD sharing (Kaipia, et al., 2017).

Supply Chain Governance

Inter-organizational governance has been studied for many years now (Poppo and Zenger, 2002; Richey et al., 2010; Narayanan and Narasimhan, 2014; xxeditors). Governance mechanisms can generally be described as a series of regulation and coordination of activities through a variety of formal and/or informal mechanisms (Williamson, 1999; Fawcett et al., 2006). Governance mechanisms in SCH management refer to the type of practices used by firms to handle their relationships with their suppliers. The choice of governance mechanisms whether arm's-length versus partnership has effects on leveraging supplier characteristics to achieve improved supplier performance (Narayanan and Narasimhan, 2014).

Governance of inter-organizational relationships can generally be contractual and/or relational (Poppo and Zenger, 2002; Richey et al., 2010; Narayanan and Narasimhan, 2014). Contractual governance is based on binding formal agreements that specify obligations and roles of exchange partners (Cao and Lumineau, 2015) and include policies, guidelines, rules, laws, norms, standards, financial and other incentives (Wathne and Heide, 2004; Bitran et al., 2007). Formal contracts provide the safeguard to protect the individual entities from bounded rationality or opportunistic behavior (Tate et al., 2010). They determine compulsory deliverables, monitoring policies, verification procedures and non-compliance penalties. Generally speaking, formal governance mechanisms have high ex-ante contractual costs and ex-post monitoring and enforcement costs (Williamson, 1985).

Relational governance on the othre hand, refers to trust-based social and moral norms in supply chain relations (Narayanan and Narasimhan, 2014; Yam and Chan, 2015). Relational governance assumes providing benefits similar to those of contractual governance in terms of controlling opportunism and facilitating adaptation (Heide and John, 1992). It is a self-reinforcing governance mechanism under which mutual understandings guide participants' interactions and adaptations (Mahapatra et al., 2010). Exchange parties learn about each other through repeated interactions and form mutual norms, values and expectations (Macneil, 1980; Chakkol et al., 2018; Zhang, et al., 2020). The use of relational forms of governance is commonly cited in the literature as viable and efficient methods for successfully governing and safeguarding inter-organizational relationships (Poppo and Zenger, 2002; Liu et al., 2009), especially in contextual settings where there is less power asymmetry (Pilbeam et al., 2012; Bonatto, et al., 2020). Contractual and relational governance mechanisms can be complementary and are therefore often used simultaneously in practice (Poppo and Zenger, 2002; Mahapatra et al. 2010; Blome, Schoenherr, and Kaesser, 2013; Cao and Lumineau, 2015).

Trust is inherent in relational governance (Liu et al., 2009; Wilson et al., 2015; Zhu et al., 2017; (Jokela & Söderman, 2017). Relational governance mechanisms are used in this study to address the issues related to BD sharing and to create a situation of "digital trust" that would allow for the expansion of BD collaboration. An investigation of the evolution of collaborative supply chain information sharing approaches related to power balance and solidarity is necessary to identify BD sharing benefits and address the risks, and are discussed in the following section.

Solidarity. In supply chain collaboration literature solidarity refers to the expectation that suppliers and buyers will generally act in a way that enhances mutual benefit, engage in bilateral problem solving, and commit to joint and coordinated action toward shared goals (Macneil, 1980; Heide and John, 1992; Bonatto et al., 2020). Solidarity emphasizes the need for some level of common appreciation, discernment, comprehension and agreement across key organizational traits, ideals, principles, and cultures (Seo et al., 2016; Frankel et al., 2002), and depends on the degree to which supply chain partners agree to work towards the achievement of a common goal (Vanathi & Swamynathan, 2014; Frankel et al., 2002; Samaddar et al., 2006). ***Goal congruence*** in supply chain management is defined as the degree to which firms believe that common goals can be achieved through multiple interactions that help collaborative members understand each other's' constraints and opportunities (Jap, 1999). It is dependent to the extent to which partners perceive their own objectives are satisfied by accomplishing the supply chain objectives (Cao and Zhang, 2011). Firms will share strategic information to achieve goals that are mutually beneficial if knowledge-sharing routines can become a source of competitive advantage (Dyer and Singh, 1998). ***Incentive alignment*** strongly affects solidarity. Successful supply chain partnerships require that each participant share gains and losses fairly and that the outcomes of the collaboration are quantifiably beneficial to all (Manthou et al., 2004; Srivastava, Srinivasan & Iyer, 2015). Incentive alignment includes determining costs, risks, and benefits (Simatupang and Sridharan, 2005; Cao and Zhang, 2011; Soosay and Hyland, 2015), and outlining a careful definition of mechanisms that share gains equitably (Lee and Whang, 2001). Solidarity includes the sharing of common goals and incentive alignment. In the absence of solidarity firms will have no motivation to participate in information sharing (Samaddar et al., 2006). Only when the alignment of goals and incentives are laid out can the full value of sharing information be captured across supply chain partners (Lee, 2004).

Perceived Fairness. Fairness is a perception that is based on the other party's actions and behaviors in a buyer-supplier relationship (Adams, 1965; Greenberg, 1987; Messick and Cook, 1983). Justice can serve as a foundation for a relationship only when both parties jointly perceive its existence (Luo, 2005). Johnson et al. (2002) and Duffy et al. (2013) suggest that perceived fairness in economic gains is important in long-term relationships; if the parties perceive that are not treated fairly when they share outcomes and incentives, it can lead to conflict and withdrawal from the relationship. The relational governance approach in supply chain management calls for an examination of the social side of justice (Liu et al., 2012). Several scholars suggested that justice be included in studying buyer–supplier relationship management (Anderson and Weitz, 1989; Dwyer et al., 1987; Frazier et al., 1988; Ireland and Webb, 2007; Narasimhan et al., 2009; (Kumar et al., 1995; Griffith et al., 2006). This view originated in the social exchange theory (Homans, 1961) and the equity theory (Adams, 1963, 1965), the first forms of which primarily emphasized the economic aspects of fairness. Extant studies focused on fairness-related gains and indicate that fairness has the ability to enhance buyer-supplier relationships due to its economic and social components (Cox et al., 2001; Anderson and Weitz, 1992; Liu et al., 2012; Luo, 2009; Zaefarian et al., 2016). In the relationship literature, fairness is referred to as the fair treatment of business partners including the fair sharing of financial terms and impartiality in decision making (Jap, 2001; Luo, 2009), that help in defining long-term orientation in a buyer-supplier relationship (Griffith et al., 2006). Addressing fairness in BD sharing between buyer and suppliers is essential in generating additional insights on effective collaborations.

In studying the structural aspect of justice (Tyler and Bies, 1990) two dimensions were identified around the concerns for equity distribution and formal procedures. These are acknowledged in the literature as the two dimensions of fairness; distributive fairness and procedural fairness. ***Distributive fairness*** refers to people’s reaction to how resources or outcomes are allocated (Liu et al., 2012). It was first proposed by Homans (1961) who emphasized that people are concerned about whether outcomes are fair rather than what the outcomes actually are. Adams (1965) then defined distributive justice as equity, suggesting that fairness exists when a person perceives that the ratio of outcomes to inputs are equal to the ratio of outcomes to inputs of the other party. Distributive fairness therefore includes the of sharing economic outcomes and equating the reward to each party's input in supplier-buyer relationships (Kumar et al., 1995). Because of the disadvantages that might be created from power imbalances or related size differences (Schleper et al., 2017), distributive fairness can include one party's willingness to take responsibility of the costs that occur in a relationship (Kumar et al., 1995) in a way that would make it acceptable to the other party perception of fairness (Mayer, 2007). ***Procedural fairness*** on the other hand lies at the junction of the economic and the social and focuses on people’s reactions to the procedures used for resolving disputes and allocating outcomes. This was introduced by Thibaut and Walker (1975) who suggested that people are often concerned about fairness of the process, and will view procedures as fair if they perceive that their party has control over the process that leads to the outcomes (Leventhal, 1980; Thibaut and Walker, 1978; Tyler and Lind, 1992). Specifically, having a say in the decision-making process with regard to sharing the actual outcome (Duffy et al., 2013; Griffith et al., 2006). Procedural fairness includes the consistency of the parties’ policies toward each other (referred to as impartiality), and the extent to which one party can challenge these policies (Kumar et al., 1995; Tyler and Lind, 1992). This procedural dimension fairness has been captured in both relational and contractual governance mechanisms. Data governance considerations are also needed to address procedural fairness perceptions.

Kumar et al. (1995) confirm that a buyer’s perceptions of distributive and procedural fairness enhance relationship quality with the supplier. Griffith et al. (2006) corroborate that both distributive and procedural fairness are equally important and create a long-term orientation for buyers. Although firms often focus on the distributive fairness and the equity of economic outcomes in a relationship, procedural fairness is critical for relational outcomes because while the nature of the distributive assessments can change, the policies and procedures are continuously observable (Kumar et al., 1995; Tyler and Lind, 1992). For the purpose of our study we will use perceived distributive fairness and perceived procedural fairness in addressing concerns with BD sharing. In dyadic exchanges, understanding both parties’ perceptions are of particular importance because the two parties need to leverage each other's capabilities and resources for achieving mutual goals (Palmatier et al., 2007; Liu et al., 2012).

The concept of “perceptual convergence within the dyad” (Anderson and Weitz, 1992, p. 29), will be used in this study.

RESEARCH METHODOLOGY

In this research, we adopt design science research (DSR) methodology. Design science method is an outcome-based research paradigm that pursues creating novel information technology (IT) artifacts (Von Alan et al., 2004). In principal, design science aims at understanding and refining the search among probable constructs and components in order to develop artifacts that are projected to solve an existing problem or challenge in their natural setting (Baskerville, 2008). DSR attempts to design and improve IT artifacts. IT artifacts could be systems, methods, algorithms, instantiations or a combination thereof. Then, enhance the functional performance of such artifacts through the continuous focus on the development process and performance evaluation (Von Alan et al., 2004). IT artifacts may vary from software applications, formal logic, and rigorous mathematical models and equations, to informal narratives and descriptions in a natural language (Von Alan et al., 2004). According to March and Smith (1995), design science outputs can be categorized in four main types, which include constructs, models, methods, and implementations. However, several research studies have extended these categories to include architecture, design principles, frameworks, instantiations, and theories (Chatterjee, 2015; Kuechler & Vaishnavi, 2019). From an abstract level, artifacts can be product or process artifacts (Gregor & Jones, 2007). Several IT artifacts have some level of abstraction, but can be potentially transformed to a more material form, like an algorithm converted to an operational model or software (Gregor & Jones, 2007). Product artifacts are usually technical or socio-technical. However, process artifacts are always socio-technical. Socio-technical artifacts are ones whose target users/humans must interact with, in order to provide their intended functions (Venable et al., 2012). At the core of the DSR is the concept of learning through the systematic building and creation of knowledge, in which the research outcomes should deliver new, innovative, true, and interesting knowledge, designs, or artifacts within the respective community of interest (Kuechler & Vaishnavi, 2019). During this creation process, the researcher must be aware of progressing both the design process and the design artifact as part of the research (Von Alan et al., 2004). In this research, the six research steps will be used to develop a complete artifact, leading to the solution (Peffers, et al., 2006) (Elragal & Hassanien, 2019). The process begins by Problem Identification and Motivation, which is has partly been explained in the previous sections. Second, to identify Objectives of a Solution, which explains how the suggested solution would work to help Information Sharing and Big Data Analytics through the different implementation phases. During the third step of the research, it would be looked into the Design and Development of the artifact to serve the purpose defined in the objective. Then the artifact is to be Demonstrated to be Evaluated and tested by examining how far the developed artifact is accepted and expected to perform in actual or real-life situation(s). Finally, the end of the research journey comes through the Communication through academic publications. It should be noted that this is an iterative process where different outputs are produced at the end of each iteration. In this research, the first iteration cycle produces the tentative design for the research project, while the following iteration cycles are used to fine-tune and enhance this tentative design, till ending with the final suggested roadmap based on all the observations during the different testing iterations of the artifact.

The Design and Evaluate Cycle

We follow the DSR guidelines explained by (Hevner, et al., 2004), in order to design, build, and evaluate the IT artifacts. That being said, the artifact in our research is a process for big data sharing. Towards that end, the process artifact that will be designed and tested in a socio-technical context. Our aim is to create knowledge that can be used to design practical big data sharing process artifact, with the necessary actions, processes, and design principles. This research is under the representation genre in DSR. We are set to evaluate in the FMCG sector in Egypt.

The objective of the artifact is to big data sharing. in order to define the problem, in-depth interviews were conducted with executives from several organizations. Selection of the interviewees was based on convenience, see Table 1. By including Salespeople, Logistics and IT from the FMCG Suppliers side, and Procurement, Purchasing and IT from the Modern Traders side. In the same time, by including multinational and local FMCG manufacturers and importers with both long and short shelf life products, and a chain of supermarkets and other chain of hypermarkets.

Table 1: In-depth Interviews: Sampling

Position	Function	Organizational Side	Organization Type
Chief Information Officer	IT & IS	Modern Trader	Local
Supply Chain Director	Supply Chain	Modern Trader	Local
Procurement Manager	Supply Chain	Modern Trader	Local
Ordering & Purchasing Manger	Supply Chain	Modern Trader	Local
IT Manager	IT & IS	Modern Trader	Local
Operation Manager	Supply Chain	FMCG Supplier	Local
Logistics & Distribution Manager	Supply Chain	FMCG Supplier	Multinational
Sales Manager	Sales	FMCG Supplier	Multinational
Sales Manager	Sales	FMCG Supplier	Local
IT Manager	IT & IS	FMCG Supplier	Local

Based on the problems identified and analysis, the design of a resolution began. The core solution component was used to resolve the raised concerns, then four different process flowcharts were developed based on implementation planning stage, to ensure readability and be clearer and more focused during the evaluation phase of design science research. Throughout the following paragraphs, each flowchart will be discussed to explain how tentative designed socio-technical artifact was carried out.

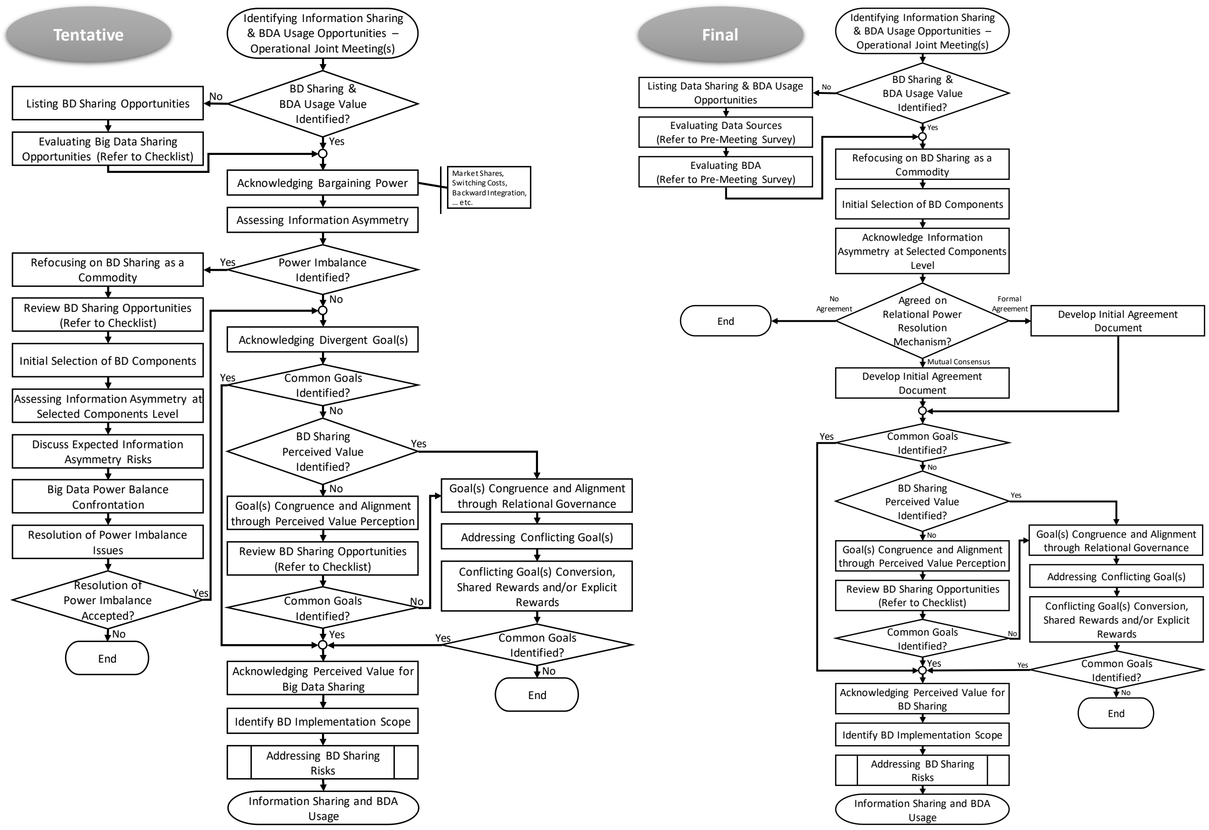


Figure 1: Process Flowcharts of First Implementation Planning Stage

Figure 1 shows the tentative and the final process flowchart of the first part that handles Power Perception and Divergent Goals, the related concerns to Inter-organizational Relationship Governance are demonstrated.

Starting with the value identification of BD Sharing and BDA Usage, by reviewing and discussing the potential opportunities. In the tentative design, the value identification was done using a list of Data Sources and BDA Models and Techniques that were identified based on literature in SCM and POM focusing on Order Replenishment Cycle. When the usage of this list was tested in real-life, users took much time to communicate and explain their opinion and outputs with others and sometimes a facilitator was needed. Therefore, it was replaced by a comprehensive survey based on the same list using MS Excel, that automatically summarize and display the results instantly in a matrix format.

This pre-meeting survey consists of two main parts; namely: Big Data Sources and Big Data Analytics. At each part, both Buyer and Supplier stakeholders individually are requested to indicate their perceived importance level of each of the listed Data Sources or Data Analytics Methods and Techniques in the four Supply Chain decision domains related to Order Replenishment Cycle. Then their responses could be collected via the developed MS Excel file, which classifies in a matrix format Data Sources and Data Analytics Methods and Techniques into 3 main categories for each of the Supply Chain decision domains; as following:

- Important Data Sources and Data Analytics Models and Techniques to both entities, which should be used as the immediate starting point of DB Sharing BDA Usage implementation.
- Somehow Important Data Sources and Data Analytics Models and Techniques to both entities, which could be reconsidered or discussed during the DB Sharing BDA Usage implementation planning meetings and/or in future enhancement stages.
- Unimportant Data Sources and Data Analytics Models and Techniques to both entities, which could be dropped in DB Sharing BDA Usage implementation.

This pre-meeting survey and sample of the results can be found in APPENDIX B: PRE-MEETING SURVEY AND SAMPLE RESULTS. Resulted matrixes facilitates the internal and external communication between the different stakeholders and entities. It could be the base of the perceived value identification and BD Scope Determination of the implementation. In the same time, it helps in handling the Power Perception and Divergent Goals as will be explained in the following paragraphs.
After that the Power Perception related concern was partially tackled targeting to have mutual consensus, by acknowledging, assessing, and confronting bargaining power and information asymmetry before and after implementation, refocusing on the data as a commodity. This was developed based on the theoretical foundations of SC Relationship Governance, using the Distributive Fairness and Procedural Fairness design principals. Real-life testing suggests adding another scenario of having an initial agreement document, and to reduce the power related confrontation by focusing only on the agreed-on Data Sources, and BDA Models and Techniques.
The next step is to handle the exiting Divergent Goals related concerns, through either Perceived Value Perception or Shared and/or Explicit Rewards. This was developed based on the theoretical foundations of SC Relationship Governance, using the Goal Congruence and Incentive Alignment design principals. Real-life testing verified the suggested initial design.
The last block in the process flowchart called “Addressing BD Sharing Risks” will be explained in detailed in the next figure. At this stage, the BD Scope Determination component should be completely covered.

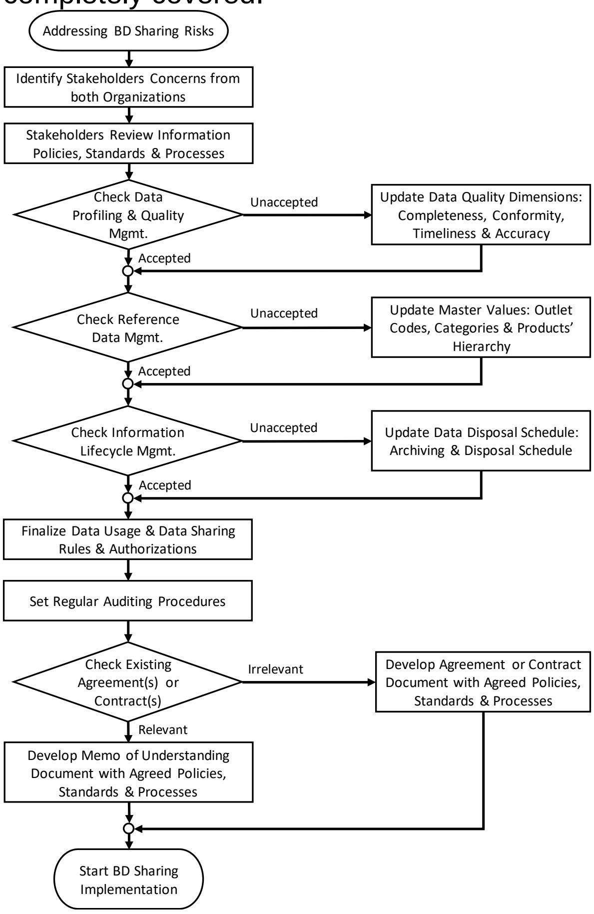


Figure 2: Process Flowcharts of Addressing BD Sharing Risks

Figure 2 shows the process flowchart of the second part that handles Addressed Risks, the last related concern to Inter-organizational Relationship Governance is demonstrated.
Starting with identifying the concerned stakeholders from both organizations, who will review the internal Information Policies, Standards, and Processes at both organizational trying to set a unified one to be used for this project. Three main checks should take place, namely: Data

Profiling and Quality Management, Reference Data Management, and Information Lifecycle Management. This was developed based on the theoretical foundations of Data Governance, focusing on BDG design principal. Then it is important to clearly set the data usage and sharing rules and authorizations, and to set a regular auditing procedures and schedule. This should be done based on the agreed and determined BD scope and the updated Information Policies, Standards, and Processes. At this stage, the data related part of the Addressed Risks component should be completely covered.

Finally, to document and formalize this agreement, a Contract or a Memo of Understanding document should be developed and signed by the authorized stakeholders stating all these agreed policies, standards, and processes, and including the auditing procedures and schedule. This was developed based on the theoretical foundations of SC Relationship Governance, using the Formal and Informa Inter-organizational Relationship Governance design principals. At this stage, the Identify Organizational and Partnership Structure and the SC relationship related part of the Addressed Risks components should be completely covered. Real-life testing verified the suggested initial design, but there was a request to add a process for collaboration objectives long-term alignment maintenance, accordingly the following separate process flowchart was developed.

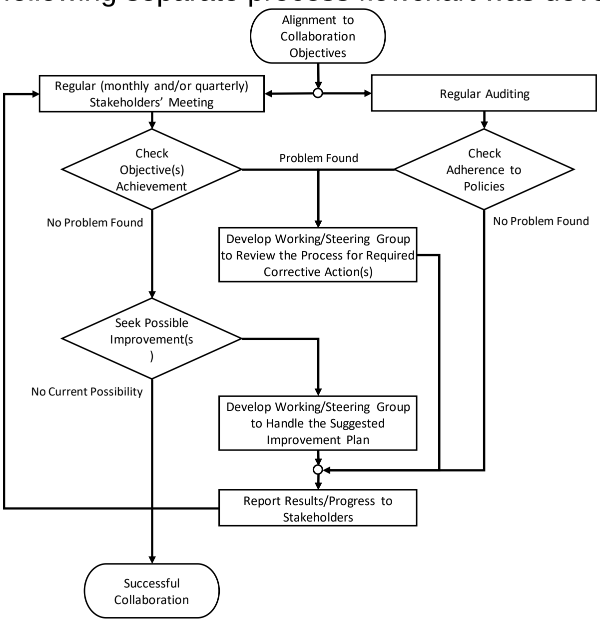


Figure 3: Process Flowchart of Collaboration Objectives Alignment Maintenance Process

Figure 3 shows the suggested process flowchart of collaboration objectives alignment maintenance process, to ensure the continuity of the collaboration relationship. Starting with collecting the output of the regular audit report and the regular stakeholders meeting. This should include reviewing the Data Profiling and Quality Management, Reference Data Management, and Information Lifecycle Management. In case of possible improvement, raised problems or required amendments, it is suggested to nominate a workforce or a committee to review the situation and develop a corrective action plan, then follow-up on its implementation after getting the approval of the stakeholders. This was developed based on the theoretical foundations of SC Relationship Governance, using again the Formal and Informal Inter-organizational Relationship Governance design principals. Real-life testing verified the suggested design.

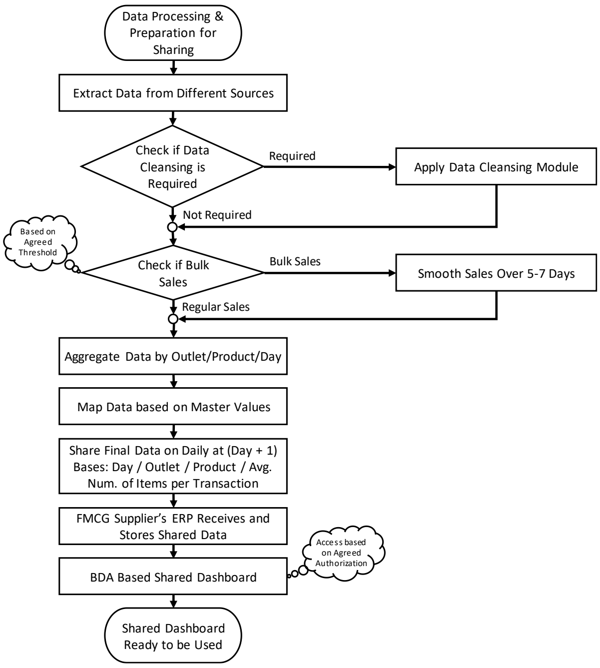


Figure 4: Process Flowchart of Data Processing & Preparation for Sharing

At this stage, the three components should be covered and fulfilled, and ready for implementation. Figure 4 illustrates the steps that are suggested to be used for data processing and preparation. Starting with extracting the data from the different sources and apply the required data cleansing if needed. Then smooth any transaction with a value that exceeds the previously agreed threshold over the following 5 to 7 days, to ensure the data privacy and reducing any potential data outliers' effect. Then transform the extracted data by applying the required aggregation, formatting, and mapping. Finally, the data are loaded or transferred to the supplier's ERP (usually at the supplier side, but other options can be considered) to be ready for analysis using the agreed BDA Techniques and Models then display the resulting outputs on a shared dashboard. This was developed based on the theoretical foundations of Data Governance, focusing on BDG design principal. Data Smoothing step was requested during real-life testing, and it was added accordingly. In the same time, Extract, Transform, and Load (ETL) procedure was adopted (not Extract, Load, and Transform (ELT)) due to some concerns raised during real-life testing related to data security and privacy complying. Based on these amendments the suggested design was verified in real-life testing.

FINDINGS

Buyer-Supplier Power Perception, Buyer-Supplier Divergent Goals, and Addressed Risks are taken care of through making use of the different Design Principals interventions used within the process of BD Sharing and BDA usage implementation. To be able to carry out these different implementation concerns and/or barriers, the suggested Socio-Technical Artifact relies on several Design Principals derived from SC Relational Governance and BD and Data Governance; namely: Goal Congruence, Incentive Alignment, Distributive Fairness, Procedural Fairness, Formal and Informal Inter-organizational Relationship Governance, and Big Data Governance. Figure 3-6 describes and explains the relation between these Design Principals the different implementation concerns and/or barriers, highlighting the effect of each Design Principal being used to resolve on the different concerns and/or barriers and reach to the point of BD Sharing and BDA Usage Opportunity identifying which is the basis of execution.

Figure 4: Relation between Design Principals and Implementation Concerns/Barriers

Artifact Evaluation

In line with the Design Science, the artifact is classified as a socio-technical process artifact. In this section, it will be explained how the artifact is going to support problem identification, solution, and verification of the solution via field evaluation. Field evaluation was compiled with three instantiations, and each instantiation was evaluated to produce an input for redesign. The third iteration was the stop as the result of the field evaluation was satisfactory with no additional expected value to the suggested designed artifact. In the previous sections, field evaluation was documented, outlining the researcher’s role and highlighting what has been done based on the field evaluation of instantiations. As the suggested designed artifact includes significant human impact on performance, the designed artifact was evaluated by potential users (as stakeholders) during repetitive in-depth interviews to verify that it can be used and implement appropriately.
Starting by interviewing around 10 participants, all of them are potential users. Then several structured individual interviews were conducted with most of them, focusing on the value of designed artifact instantiations, they were asked about their feedback of the designed artifact, if it gives sufficient information to use it as a model for their own efforts in reducing Order Replenishment Cycle planning. Results of each interview provided a feedback towards redesigning the next instantiations, and to determine which of the developed Design Principals were most valuable for them. Finally, their feedback was used to developed the previously communicated final process flowcharts.

Generalizability

Ideally, Design Science must involve the possibility to generalize the output artifact. In reality, the BD Sharing and BDA Usage implementation process flowchart as the artifact resulting from this Design Science research is a collection of ingredients serving different purposes,

Big Data Sharing and Governance in Supply Chains

therefore, the artifact could be generalized. Four levels of generalization could be discussed, and check how they apply to the suggested artifact solution (Elragal & Hassanien, 2019):

1) Problem-Related Generalizations: Suggested artifact solution is designed to address problem-related to BD Sharing and BDA Usage implementation. However, it could also apply to similar situations without either BD or Information Sharing. For instance, most of the suggested design principles could also apply to Information Sharing of structured data and traditional analytics, also, it is possible to apply most of the suggested design principles for BDA Usage implementation within a single organization.
2) Solution-Related Generalizations: Some of the suggested artifact components for BD Sharing and BDA Usage implementation could be used as part of other solutions, such as the BD Scope Determination component, and the same goes for the BDG Policies and Standards component.
3) Design Principles: The knowledge gained during Design and Implementation of this artifact solution designed to address BD Sharing and BDA Usage implementation could serve as Design Principles, for future researchers and practitioners to use. Below is the list of Design Principles which have been used during the artifact design:
    - Goal Congruence: it is important to understand each other’s constraints and opportunities and to take into account mutually beneficial goals and overall competitive advantage.
    - Incentive Alignment: it is important to carefully outline and identify mechanisms that ensure sharing gains, losses, and risks equitably, as it strongly affects solidarity.
    - Distribution Fairness: it is important to create a long-term orientation for both entities in the relationship and to enhance relationship quality, as it refers to the perception of how resources or outcomes are allocated. This includes willingness to take responsibility of the costs that occur in a relationship by one party in a way that would make it acceptable to the other parties.
    - Procedural Fairness: it is also important to create a long-term orientation for both entities in the relationship and to enhance relationship quality, as it refers to procedures used in decision-making process, resolving disputes, and allocating outcomes, which has been captured in both Inter-organizational Relationship Governance and Big Data Governance.
    - Formal and Informal Inter-organizational Relationship Governance: it is important to create a situation of “Digital Trust" and to achieve collaboration performance, as it refers to series of regulation and practices through a variety of formal and/or informal (or contractual and/or relational) mechanisms. Contractual and Relational Governance mechanisms are complementary and are often used simultaneously in practice.
    - Big Data Governance: it is important to ensure data reliability, data security, and data privacy, as it refers to policies, rules, standards and procedures related to data gathering, quality, storage, management, usage, lifecycle, privacy, security, and regulatory compliance. Inter-organizational BDG scope focuses: Data Integration and Usage Policies, Data Exchange Standards, Processes for Interaction and Collaboration, Data Sharing Agreements, … etc.
4) Feedback to Design Theory: It determines the generalizability range for the suggested Design Principles, and the all the above mentioned efforts. This may include the possible enhancement of the usage of contributing theories, as a complete design theory may not emerge from a single Design Science research attempt.

CONCLUSION

Importance of both Information Sharing and BDA in SCM and POM, has been proven. The artifact has been formulated to facilitate the implementation planning. The suggested BD Sharing and BDA Usage designed artifact formulates a socio-technical self-moderated process, handling the complications of the technical and social implementation planning activities and aspects by tackling the Power Perception, Divergent Goals, and Relational and Data Addressed Risks. The artifact works by guiding the stakeholders to resolve these raised concerns and/or problems throughout the different implementation activities. Such concerns and/or problems arising from technical, business needs, and previous Behavioral Science researches. By deeply analyzing these concerns and/or problems, the artifact could tackle and resolve them.

This research work discussed and suggested process flowcharts that work on addressing the BD Sharing and BDA Usage implementation challenges by building on and using the theoretical foundation of SCM and BDG. The six-phased Design Science model has been adopted to help pave the research path. This could be achieved through several iterations of the design and development of the Design Science artifact. Modularity of the designed artifact also supports the generality of the design and how it would be able to be applied in different contexts.

Limitations of the Study and Suggested Future Work of Research

Selection of proper BDA models and techniques, especially the predictive and prescriptive techniques, are somehow dependent on data sources, reliability, and the targeted organization itself and its objectives. Implementation and reliability test needs access to real-life data, and needs time to validate against reality, which is somehow non-business-related topic. Therefore, this research will not focus on this angle, and will only use descriptive analytics, keeping it open as a further point of research.

Implementation cost calculations depends on the additional IT infrastructure required, and expected benefit estimations needs to be measured against current performance in Operational Efficiency, Customer Experience, and Research and Development. Simulation can be used for understanding the performance and identifying the benefits of supply chain collaboration initiatives, without investing large amounts of money in establishing the actual collaboration (Ramanathan, 2014). In the same time, simulation and system dynamics can provide an understanding to the effect of changes made in demand planning by using Big Data Analytics (Hofmann, 2015). Cost/Benefit analysis and split between entities for the suggested collaboration, using simulation needs an access to real-life data, therefore, this research will not focus on this angle, keeping it open as a further point of research.

It was earlier explained that the designed artifact seeks to promote BD Sharing and BDA Usage in Dyadic Relationship but relaying on the currently available Database Management Systems either on-cloud or on-premises. In future, studying how business process integration would be feasible, especially in the light of IoE and blockchain technologies; that is: to study the application of state-of-the-art technologies as well as business related-organizational impacts.

Among the limitations, this research study has not focus on the problems of end users interacting with the proposed BD Sharing and BDA Usage. In the same time, given the expected impact of the BD Sharing and BDA Usage designed artifact, the way organizations are going to engage may significantly change due to the promoted level of communication between stakeholders. It should be noted that future research may also include testing the artifact among end users in various organizational contexts or in different sectors, which may lead to add new Design Principles to the ones already identified earlier on. That said, future scholarly publications should illustrate and discuss observed findings from the design, development, demonstration, and evaluation phases by assessing use-case scenarios over several real-life cases, while also adopting the Action Design research methods to ensure organizational relevance.

Finally, because of human agency and lack of comparable mechanisms, exact behavior and performance of the suggested solution cannot be predicted with full certainty, as applying Design Science to SCM and POM system may present some issues that makes it difficult to generalize it.

The effect of human agency is significant in SCM and POM systems, as human behavior is strongly affected by personal relations and context. Therefore, performance of system behavior cannot be predicted with high certainty, making the solution design's pragmatic validity somehow difficult to be established. In the same time, weak mechanisms of social domain enable prediction of social behavior, but with less certainty than with strong mechanisms. On the other hand, if human agency impact on a SCM and POM system is limited, structured data gathering, and mathematical modeling can be powerful approaches. If variation in the context caused a significant variation in the system performance, then the design solution should be further generalized by testing it in different contexts of the intended application domain.

REFERENCES

Abraham, R., Schneider, J. & Brocke, J., 2019. Data Governance: A Conceptual Framework, Structured Review, and Research Agenda. *International Journal of Information Management,* 24 July, Volume 49, pp. 424-438.

Adams, J. S., 1965. Inequity in Social Exchange. In: L. Berkowitz, ed. *Advances in Experimental Social Psychology.* New York(New York): ScienceDirect - Elsevier, pp. 267-299.

Addo-Tenkorang, R. & Helo, P. T., 2016. Big Data Applications in Operations/Supply Chain Management - A Literature Review. *Computers & Industrial Engineering,* Volume 101, pp. 528-543.

Al-Badi, A., Tarhini, A. & Khan, A. I., 2018. *Exploring Big Data Governance Frameworks.* Leuven, Belgium, ScienceDirect - Elsevier, pp. 271-277.

Anderson, E. & Weitz, B., 1989. Determinants of Continuity in Conventional Industrial Channel Dyads. *Marketing Science,* 01 November, 8(4), pp. 291-380.

Anderson, E. & Weitz, B., 1992. The Use of Pledges to Build and Sustain Commitment in Distribution Channels. *Journal of Marketing Research,* 01 February, 29(1), pp. 18-34.

Angeles, R. & Nath, R., 2001. Partner Congruene in Electronic Data Interchange (EDI) - Enabled Relationships. *Journal of Business Logistics,* 17 November, 22(2), pp. 109-127.

Bazyar, A. et al., 2013. Linking Power, Risk, and Governance: a Survey Research in New Product Development Relationships. *Journal of Business & Industrial Marketing,* 07 June, 28(5), pp. 371-382.

Biswas, S. & Sen, J., 2016. A Proposed Architecture for Big Data Driven Supply Chain Analytics. *The IUP Journal of Supply Chain Management,* September, XIII(3), pp. 7-33.

Bitran, G. R., Gurumurthi, S. & Sam, S. L., 2007. The Need for Third-Party Coordination in Supply Chain Governance. *MIT Sloan management review,* Spring, 48(3), pp. 30-37.

Blazquez, D. & Domenech, J., 2018. Big Data Sources and Methods for Social and Economic Analyses. *Technological Forecasting & Social Change,* May, Volume 130, pp. 99-113.

Blome, C., Schoenherr, T. & Kaes, M., 2013. Ambidextrous Governance in Supply Chains: The Impact on Innovation and Cost Performance. *Journal of Supply Chain Management,* 21 October, 49(4), pp. 59-80.

Bonatto, F., Martins de Resende, L. M. & Pontes, J., 2020. Relational Governance in Supply Chain: a Systematic Literature Review. *Benchmarking: An International Journal,* June June, 27(6), pp. 1711-1741.

Cagliano, R., Caniato, F. & Spina, G., 2003. E-Business Strategy How Companies are Shaping their Supply Chain Through the Internet. *International Journal of Operations & Production Management,* 01 October, 23(10), pp. 1142-1162.

Cao, M. & Zhang, Q., 2011. Supply Chain Collaboration: Impact on Collaborative Advantage and Firm Performance. *Journal of Operations Management,* March, 29(3), pp. 163-180.

Cao, Z. & Lumineau, F., 2015. Revisiting the Interplay between Contractual and Relational Governance: A Qualitative and Meta-Analytic Investigation. *Journal of Operations Management,* January, Volume 33-34, pp. 15-42.

Chakkol, M., Selviaridis, K. & Finne, M., 2018. The Governance of Collaboration in Complex Projects. *International Journal of Operations & Production Management,* 03 April, 38(4), pp. 997-1019.

Chandra, C. & Grabis, J., 2007. Scope of Supply Chain Configuration Problem. In: *Supply Chain Configuration - Concepts, Solutions, and Applications.* Boston(Massachusetts): Springer, pp. 17-45.

Chandrasekaran, A., de Treville, S. & Browning, T., 2020. Editorial: Intervention-Based Research (IBR) - What, Where, and How to Use it in Operations Management. *Journal of Operations Management,* 03 June, 66(4), pp. 370-378.

Chatfield, D. C., Kim, J. G., Harrison, T. P. & Hayya, J. C., 2004. The Bullwhip Effect - Impact of Stochastic Lead Time, Information Quality, and Information Sharing: A Simulation Study. *Production and Operations Management Society,* 01 December, 13(4), pp. 340-353.

Chehbi-Gamoura, S., Derrouiche, R., Damand, D. & Barth, M., 2020. Insights from Big Data Analytics in Supply Chain Management: an All-Inclusive Literature Review Using the SCOR Model. *Production Planning & Control,* 31(5), pp. 355-382.

Cox, A., Sanderson, J. & Watson, G., 2001. *Power Regimes: A New Perspective on Managing in Supply Chains and Networks.* Jönköping, Sweden, IPSERA, pp. 215-227.

Dejonckheere, J., Disney, S. M., Lambrecht, M. R. & Towill, D. R., 2003. Measuring and Avoiding the Bullwhip Effect: A Control Theoretic Approach. *European Journal of Operational Research,* 16 June, 147(3), pp. 567-590.

Dev, N. K., Shankar, R., Gupta, R. & Dong, J., 2018. Multi-Criteria Evaluation of Real-Time Key Performance Indicators of Supply Chain with Consideration of Big Data Architecture. *Computers & Industrial Engineering,* 10 April, Volume 128, pp. 1076-1087.

Disney, S. M., Naim, M. M. & Towill, D. R., 1997. Dynamic Simulation Modelling for Lean Logistics. *International Journal of Physical Distribution & Logistics Management,* 27(3/4), pp. 174-196.

Duffy, R. et al., 2013. Engaging Suppliers in CRM: the Role of Justice in Buyer–Supplier Relationships. *International Journal of Information Management,* February, 33(1), pp. 20-27.

Dwyer, F. R., Schurr, P. H. & Oh, S., 1987. Developing Buyer-Seller Relationships. *Journal of Marketing,* 09 April, 51(2), pp. 11-27.

Dyer, J. H. & Singh, H., 1998. The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage. *Academy of Management Review,* 01 October, 23(4), pp. 660-679.

Elragal, A. & Haddara, M., 2019. Design Science Research: Evaluation in the Lens of Big Data Analytics. *Systems,* 28 May, 7(2), pp. 1-8.

Elragal, A. & Hassanien, H. E.-D., 2019. Augmenting Advanced Analytics into Enterprise Systems: A Focus on Post-Implementation Activities. *Systems,* 21 June, 7(2), pp. 1-24.

Fawcett, S. E., Ogden, J. A., Magnan, G. M. & Cooper, M. B., 2006. Organizational Commitment and Governance for Supply Chain Success. *International Journal of Physical Distribution & Logistics Management,* 01 January, 36(1), pp. 22-35.

Feki, M., Boughzala, I. & Wamba, S. F., 2016. *Big Data Analytics-Enabled Supply Chain Transformation: A Literature Review.* Koloa, Hawaii, USA, IEEE Computer Society, pp. 1123-1132.

Firican, G., 2017. *Adapting Data Governance to Big Data.* [Online] Available at: https://tdwi.org/articles/2017/04/14/adapting-data-governance-to-big-data.aspx [Accessed 28 May 2020].

Firican, G., 2017. *The 10 Vs of Big Data.* [Online] Available at: https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx [Accessed 15 August 2018].

Frazier, G. L., Spekman, R. E. & O'Neal, C. R., 1988. Just-In-Time Exchange Relationships in Industrial Markets. *Journal of Marketing,* 01 October, 52(4), pp. 52-67.

Goffin, K., Lemke, F. & Szwejczewski, M., 2006. An Exploratory Study of ‘Close’ Supplier-Manufacturer Relationships. *Journal of Operations Management,* January, 24(2), pp. 189-209.

Grean, M. & Shaw, M. J., 2002. Supply-Chain Integration through Information Sharing: Channel Partnership between Wal-Mart and Procter & Gamble. In: *E-Business Management - Integration of Web Technologies with Business Models.* Boston(Massachusetts): Springer, pp. 155-171.

Greenberg, J., 1987. A Taxonomy of Organizational Justice Theories. *Academy of Management Review,* 01 January, 12(1), pp. 9-22.

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,* 01 January, 24(2), pp. 85-98.

Hartmann, P. M., Zaki, M., Feldmann, N. & Neely, A., 2016. Big Data for Big Business - A Taxonomy of Data-Driven Business Models Used by Start-up Firms. *International Journal of Operations & Production Management,* 03 October, 36(10), pp. 1382-1406.

Heide, J. B. & John, G., 1992. Do Norms Matter in Marketing Relationships?. *Journal of Marketing,* 01 April, 56(2), pp. 32-44.

Hermann, M., Pentek, T. & Otto, B., 2016. *Design Principles for Industrie 4.0 Scenarios.* Koloa, Hawaii, USA, IEEE Computer Society, pp. 3928-3937.

Hevner, A. R., March, S. T., Park, J. & Ram, S., 2004. Design Science in Information Systems Research. *Management Information Systems Quarterly,* March, 28(1), pp. 75-105.

Hofmann, E., 2015. Big Data and Supply Chain Decisions: the Impact of Volume, Variety and Velocity Properties on the Bullwhip Effect. *International Journal of Production Research,* 10 July, 55(17), pp. 5108-5126.

Homans, G. C., 1961. The Humanities and the Social Sciences. *American Behavioral Science,* 01 April, 4(8), pp. 3-6.

Horwitz, L., 2019. *The Future of IoT Miniguide: The Burgeoning IoT Market Continues.* [Online] Available at: https://www.cisco.com/c/en/us/solutions/internet-of-things/future-of-iot.html [Accessed 28 May 2020].

Huan, S. H., Sheoran, S. K. & Wang, G., 2004. A Review and Analysis of Supply Chain Operations Reference (SCOR) Model. *Supply Chain Management: An International Journal,* 01 February , 9(1), pp. 23-29.

Ireland, R. D. & Webb, J. W., 2007. A Multi-Theoretic Perspective on Trust and Power in Strategic Supply Chains. *Journal of Operations Management,* March, 25(2), pp. 482-497.

Jamehshooran, B. G., Shaharoun, A. M. & Haron, H. N., 2015. Assessing Supply Chain Performance through Applying the SCOR Model. *International Journal of Supply Chain Management,* March, 4(1), pp. 1-11.

Janssen, M., Van Der Voort, H. & Wahyudi, A., 2017. Factors Influencing Big Data Decision-Making Quality. *Journal of Business Research,* January, Volume 70, pp. 338-345.

Jap, S. D., 2001. “Pie Sharing” in Complex Collaboration Contexts. *Journal of Marketing Research,* 01 February, 38(1), pp. 86-99.

Johnson, J. P., Korsgaard, M. A. & Sapienza, H. J., 2002. Perceived Fairness, Decision Control, and Commitment in International Joint Venture Management Teams. *Strategic Management Journal,* 19 September, 23(12), pp. 1141-1160.

Jokela, P. & Söderman, A., 2017. Re-Examining the Link Between Fairness and Commitment in Buyer-Supplier Relationship. *Journal of Purchasing and Supply Management,* 22 September, 23(4), p. 268–279.

Kaipia, R., Holmström, J., Småros, J. & Rajala, R., 2017. Information Sharing for Sales and Operations Planning: Contextualized Solutions and Mechanisms. *Journal of Operations Management,* 29 April, 52(1), pp. 15-29.

Kumar, N., Scheer, L. K. & Steenkamp, J.-B. E. M., 1995. The Effects of Perceived Interdependence on Dealer Attitudes. *Journal of Marketing Research,* 01 August, 32(3), pp. 348-356.

Lee, H. L., Padmanabhan, V. & Whang, S., 1997. The Bullwhip Effect in Supply Chains. *MIT Sloan Management Review*, 15 April, 38(3), pp. 93-102.

Lee, H. L. & Whang, S., 2001. Winning the Last Mile of E-Commerce. *Standford Global Supply Chain Management Forum,* November, Volume 2, pp. 1-20.

Leventhal, H., 1980. *Toward a Comprehensive Theory of Emotion.* 13 ed. Madison(Wisconsin): Academic Press.

Liao, Y., Deschamps, F., Loures, E. d. F. R. & Ramos, L. F. P., 2017. Past, Present and Future of Industry 4.0 - A Systematic Literature Review and Research Agenda Proposal. *International Journal of Production Research,* 28 March, 55(12), pp. 3609-3629.

Liu, Y., Huang, Y., Luo, Y. & Zhao, Y., 2012. How Does Justice Matter in Achieving Buyer–Supplier Relationship Performance?. *Journal of Operations Management,* July, 30(5), pp. 355-367.

Lummus, R. R. & Vokurka, R. J., 1999. Defining Supply Chain Management: A Historical Perspective and Practical Guidelines. *Industrial Management & Data Systems,* 01 February, 99(1), pp. 11-17.

MacCarthy, B. L. et al., 2016. Supply Chain Evolution - Theory, Concepts and Science. *International Journal of Operations & Production,* 05 December, 36(12), pp. 1696-1718.

Macneil, I. R., 1980. Power, Contract, and the Economic Model. *Journal of Economic Issues,* December, 14(4), pp. 909-923.

Malik, P., 2013. Governing Big Data: Principles and Practices. *IBM Journal of Research and Development,* 17 May, 57(3/4), pp. 1-13.

Manthou, V., Vlachopoulou, M. & Folinas, D., 2004. Virtual e-Chain (VeC) Model for Supply Chain Collaboration. *International Journal of Production Economics,* 18 February, 87(3), pp. 241-250.

Merkuryeva, G., Napalkova, L. & Vecherinska, O., 2009. *Simulation-Based Analysis and Optimisation of Planning Policies over the Product Life Cycle within the Entire Supply Chain.* Moscow, Russia, Elsevier - ScienceDirect , pp. 576-581.

Messick, D. & Cook, K., 1983. *Equity Theory: Psychological and Sociological Perspectives.* New York(New York): Praeger Publishers.

Narasimhan, R. et al., 2009. Lock-in Situations in Supply Chains: A Social Exchange Theoretic Study of Sourcing Arrangements in Buyer-Supplier Relationships. *Journal of Operations Management,* October, 27(5), pp. 374-389.

Narayanan, S. & Narasimhan, R., 2014. Governance Choice, Sourcing Relationship Characteristics, and Relationship Performance. *Decision Sciences,* 25 August, 45(4), pp. 717-751.

Nargundkar, Nargundkar, S. & Daley, M., 2006. Inter-Organizational Information Sharing: The Role of Supply Network Configuration and Partner Goal Congruence. *European Journal of Operational Research,* 16 October, 174(2), pp. 744-765.

Neumann, A., n.d. *7 Areas of Supply Chain.* [Online] Available at: https://www.supplychain4me.com/7-areas-of-supply-chain [Accessed 24 August 2018].

Nguyen, T. et al., 2017. Big Data Analytics in Supply Chain Management: A State-of-the-Art Literature Review. *Computers and Operations Research,* 08 July, Volume 98, pp. 254-264.

Nitsche, B. & Straube, F., 2020. *Efficiently Managing Supply Chain Volatility - A Management Framework for the Manufacturing Industry.* Shanghai, China, ScienceDirect - Elsevier, pp. 320-327.

Oliva, R., 2019. Intervention as a Research Strategy. *Journal of Operations Management,* 17 October, 65(7), pp. 710-724.

Palmatier, R. W., Dant, R. P. & Grewal, D., 2007. A Comparative Longitudinal Analysis of Theoretical Perspectives of Interorganizational Relationship Performance. *Journal of Marketing,* 01 October, 71(4), pp. 172-194.

Peffers, K. et al., 2006. *The Design Science Research Process: A Model for Producing and Presenting Information Systems Research.* Claremont, California, USA, Claremont Graduate University, pp. 83-106.

Phillips-Wren, G., Iyer, L. S., Kulkarni, U. & Ariyachandra, T., 2015. Business Analytics in the Context of Big Data: A Roadmap for Research. *Communications of the Association for Information Systems,* August, 37(1), pp. 448-472.

Poppo, L. & Zenger, T., 2002. Do Formal Contracts and Relational Governance Function as Substitutes or Complements?. *Strategic Management Journal,* 01 May, 23(8), pp. 707-725.

Ramanathan, U., 2014. Performance of Supply Chain Collaboration - A Simulation Study. *Expert Systems with Applications,* January, 41(1), pp. 210-220.

Richey, R. G., Roath, A. S., Whipple, J. M. & Fawcett, S. E., 2010. A Comparative Longitudinal Analysis of Theoretical Perspectives of Interorganizational Relationship Performance. *Journal of Business Logistics,* Spring, 31(1), pp. 237-256.

Ross, D. f., 1998. *Competing Through Supply Chain Management.* 1 ed. Boston(Massachusetts): Springer.

Rozados, I. V. & Tjahjono, B., 2014. *Big Data Analytics in Supply Chain Management: Trends and Related Research.* Bali, Indonesia, s.n., pp. 1096-1107.

Russom, P., 2016. *As with All Data, Big Data Needs Governance.* [Online] Available at: https://tdwi.org/articles/2016/12/02/big-data-needs-governance.aspx [Accessed 28 May 2020].

Samaddar, S., Nargundkar, S. & Daley, M., 2006. Inter-Organizational Information Sharing: The Role of Supply Network Configuration and Partner Goal Congruence. *European Journal of Operational Research,* 16 October, 174(2), pp. 744-765.

Schleper, M. C., Blome, C. & Wuttke, D. A., 2017. The Dark Side of Buyer Power: Supplier Exploitation and the Role of Ethical Climates. *Journal of Business Ethics,* January, Volume 140, pp. 97-114.

Shao, G., Jain, S. & Shin, S.-J., 2014. *Data Analytics Using Simulation for Smart Manufacturing.* Savannah, Georgia, USA, IEEE Xplore, pp. 2192-2203.

Sheu, C., Yen, H. R. & Chae, B., 2006. Determinants of Supplier-Retailer Collaboration: Evidence from an International Study. *International Journal of Operations & Production Management,* 01 January, 26(1), pp. 24-49.

Siddiqa, A. et al., 2016. A Survey of Big Data Management: Taxonomy and State-of-the-Art. *Journal of Network and Computer Applications,* 22 April, Volume 71, pp. 151-166.

Simatupang, T. M. & Sridharan, R., 2005. The Collaboration Index: a Measure for Supply Chain Collaboration. *International Journal of Physical Distribution & Logistics Management,* 01 January, 35(1), pp. 44-62.

Soosay, C. A. & Hyland, P., 2015. A Decade of Supply Chain Collaboration and Directions for Future Research. *Supply Chain Management: An International Journal,* 14 September, 20(6), pp. 613-630.

Souza, G. C., 2014. Supply Chain Analytics. *Business Horizons,* 28 July, 57(5), pp. 595-605.

Srivastava, P., Srinivasan, M. & Iyer, K. N. S., 2015. Relational Resource Antecedents and Operational Outcome of Supply Chain Collaboration: The Role of Environmental Turbulence. *Transportation Journal,* Spring, 54(2), pp. 240-274.

Tan, K. C., Lyman, S. B. & Wisner, J. D., 2002. Supply Chain Management: A Strategic Perspective. *International Journal of Operations & Production Management,* 01 June , 22(6), pp. 614-631.

Tate, W. L. et al., 2010. An Agency Theory Perspective on the Purchase of Marketing Services. *Industrial Marketing Management,* July, 39(5), pp. 806-819.

Thibaut, J. & Walker, L., 1978. A Theory of Procedure. *California Law Review,* 66(3), pp. 541-566.

Tuten, T. L. & Urban, D. J., 2001. An Expanded Model of Business-to-Business Partnership Formation and Success. *Industrial Marketing Management,* February, 30(2), pp. 149-164.

Tyler, T. R. & Lind, E. A., 1992. A Relational Model of Authority in Groups. *Advances in Experimental Social Psychology,* Volume 25, pp. 115-191.

van Aken, J., Chandrasekaran, A. & Halman, J., 2016. Conducting and Publishing Design Science Research Inaugural Essay of the Design Science Department of the Journal of Operations Management. *Journal of Operations Management,* 01 July, 47-48(1), pp. 1-8.

Vanathi, R. & Swamynathan, R., 2014. Competitive Advantage Through Supply Chain Collaboration – An Empirical Study of the Indian Textile Industry. *Fibres & Textiles in Eastern Europe,* 22(4 (106)), pp. 8-13.

Waller, M. A. & Fawcett, S. E., 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics,* 11 June, 34(2), pp. 77-84.

Wang, E. T. G. & Wei, H.-L., 2007. Interorganizational Governance Value Creation: Coordinating for Information Visibility and Flexibility in Supply Chain. *Decision Sciences - A Journal of the Decision Sciences Institute,* 07 December, 38(4), pp. 647-674.

Wathne, K. H. & Heide, J. B., 2004. Relationship Governance in a Supply Chain Network. *Journal of Marketing,* 01 January, 68(1), pp. 73-89.

Westbrook, R., 1995. Action Research: A New Paradigm for Research in Production and Operations Management. *International Journal of Operations & Production Management,* 1 December, 15(12), pp. 6-20.

Williamson, O. E., 1985. *The Economic Institutions of Capitalism. Firms, Markets, Relational Contracting.* New York(New York): Free Press.

Williamson, O. E., 1999. Strategy Research: Governance and Competence Perspectives. *Strategic Management Journal,* 25 November, 20(12), pp. 1087-1108.

Witkowski, K., 2017. *Internet of Things, Big Data, Industry 4.0 - Innovative Solutions in Logistics and Supply Chains Management.* Bialystok, Poland, ScienceDirect - Elsevier, pp. 763-769.

Yam, R. C. M. & Chan, C., 2015. Knowledge Sharing, Commitment and Opportunism in New product Development. *International Journal of Operations & Production Management,* 06 July, 35(7), pp. 1056-1074.

Yan, J., Li, K. & Qiu, D., 2007. *A SCOR-Based Method for Supply Chain Process Reengineering with Applications in Chinese Automotive Industry.* Keystone, Colorado, USA, AIS Electronic Library, p. Paper 226.

Yebenes, J. & Zorrilla, M., 2019. *Towards a Data Governance Framework for Third Generation Platforms.* Leuven, Belgium, ScienceDirect - Elsevier, p. 614–621.

Yılmaz, H., Çemberci, M. & Uca, N., 2016. The Role of Collaborative Advantage for Analyzing the Effect of Supply Chain Collaboration on Firm Performance. *International Journal of Commerce and Finance,* 26 May, 2(1), pp. 157-168.

Zaefarian, G., Najafi-Tavani, Z., Henneberg, S. C. & Naudé, P., 2016. Do Supplier Perceptions of Buyer Fairness Lead to Supplier Sales Growth?. *Industrial Marketing Management Journal,* February, Volume 53, pp. 160-171.

Zhang, Q., Jin, J. L. & Yang, D., 2020. How to Enhance Supplier Performance in China: Interplay of Contracts, Relational Governance and Legal Development. *International Journal of Operations & Production Management,* 27 August, 40(6), pp. 777-808.

Zhu, Q., Krikke, H. & Caniëls, M. C., 2017. Twin-Objective Supply Chain Collaboration to Cope with Rare but High Impact Disruptions Whilst Improving Performance. *The International Journal of Logistics Management,* 08 May, 28(2), pp. 488-507.

51st Annual Conference Of The Decision Sciences Institute

APPENDIX A: In-Depth Interview Discussion Guideline

Name: ………………………………………………………………………………

Company: ………………………………………………………………………………

Position: ………………………………………………………………………………

Introduction:

Information Sharing specially in the era of Big Data can be the core of a Dyadic Relationship – between FMCG Manufacturer and Retailer. Governance of this relationship, shared data, and different generated reports is the focus of this study.

Supply Chain Management Decisions:

In a Dyadic Relationship between FMCG Manufacturer and Retailer:

1. What are the most important problems related to Supply Chain Management (Replenishment Cycle) that both partners should coordinate with each other?
[Probe to Cover the Different Functions: Inventory Management, Warehousing & Logistics]

Data Driven Decision Making:

For each of the problems listed in the previous section:

2. Is data used to solve this problem?
3. Are analytics used on the top of data to solve this problem?
[Try to See Report Samples]
4. What problems are faced in the current decision-making process?
[Probe to Get as Many Problems as Possible: Accuracy, Inconsistency, Delay, … etc.]
5. What are the main Data, Information, Knowledge, Models, Analytics that can be used, utilized and shared by both organizations to improve this decision? What is the expected output of each of these Data Analytics?
[Try to Cover the Different Types: Descriptive, Predictive & Prescriptive]

Data Governance:

For each of the Data Analytics suggested in the previous section to improve certain decision:

6. What are the concerns related to data governance, privacy, security, and quality for each of these data sources? What tools can be used?
7. What are the main issues in data selection and organization? What tools can be used?

Relationship Governance:

In Data and Data Analytics based Dyadic Relationship between FMCG Manufacturer and Retailer:

1. Who are the concerned stakeholders from both organizations?
2. How should be the structure of this Dyadic Relationship?
3. How this Dyadic Relationship should be monitored to ensure its success and continuity?
   [Probe to Ensure Covering: Fairness & Commitment]

Suggestions and Other:

4. Is there anything that would be added? Any comments to add?

APPENDIX B: PRE-MEETING SURVEY AND SAMPLE RESULTS

Pre-Meeting Survey - Part 1: Data Sources

From your point of view as a **Buyer** - focusing on Order Replenishment Cycle, please indicate the perceived Importance (1 Totally Disagree - 2 Neutral - 3 Totally Agree) of sharing each of the following Data Sources between you and **your Suplier** in each of the four listed Supply Chain decision domains…

Supplier's Data Sources[(1)]	Supply Chain Function[(1)]	Data Source Description[(2)]	Perceived Important to be Shared (1 Totally Disagree - 2 Neutral - 3 Totally Agree)			
			Demand Forecasting	Inventory Replenishment Planning	Shipment Consolidation & Distribution	Sales Promotions Planning
Structured Data Sources						
ERP Transaction Data	Ordering, Planning & CRM	Transactions' information recorded on ERP (Enterprise Resource Planning)	2	2	3	2
Product Traceability and Monitoring System	Warehousing	System that follows a product through all phases of manufacturing, warehousing and distribution process	2	1	2	3
Barcode, RFID and QR Code System	Warehousing	System that automatically identify and track Barcode, RFID (Radio Frequency Identification), and QR Code (Quick Response Code) tags attached to different objects	3	3	3	2
Sales History	Ordering, Planning & CRM	Summary data of the sales of a product for a given time period (months or years)	1	3	3	3
Point-of-Sales System	Ordering, Planning & CRM	Terminals and Systems where customers execute their payment for goods or services in physical store, or virtual sales point such as a computer or mobile electronic device	3	1	2	1
Demand Forecasts	Ordering, Planning & CRM	Demand prediction of a product in the future (months or years)	2	2	3	2
Delivery Times and Terms	Procurement & SRM	Agreed timing and/or rate of delivery between a buyer and a seller, for products purchased for a future delivery period	2	2	1	3
In-transit Inventory	Warehousing	Inventory that have shipped from the seller, but not yet received by the buyer	2	2	2	2
Semi-Structured Data Sources						
EDI Purchase Orders and Invoices	Procurement & SRM	Electronic version of Invoice sent via EDI (Electronic Data Interchange), that is typically sent in response to an electronic Purchase Order including prices and quantities	1	3	1	2
Customer Surveys	Ordering, Planning & CRM	Customer answers to questions that are used to analyze whether or not changes need to be made in business operations to increase customers' satisfaction and loyalty	1	1	3	2
Suppliers Intelligence	Ordering, Planning & CRM	Ethical monitoring of strategic suppliers (e.g. suppliers' pricing, other customers, capacity, financials, … etc..)	2	3	3	3
Competitive Intelligence	Ordering, Planning & CRM	Ethical monitoring of main competitors, main customers, and any aspect of the environment needed (e.g. competitors' pricing, capacity, financials, … etc.)	1	2	3	3
Market Intelligence	Ordering, Planning & CRM	Ethical monitoring of strategic suppliers, main competitors, main customers, and any aspect of the environment needed (e.g. competitors' pricing, capacity, financials, … etc.)	1	3	2	2
Unstructured Data Sources						
Intelligent Transport Systems (ITS)	Logistics	Application of sensing, analysis, control and communications technologies to ground transportation in order to improve safety, mobility and efficiency	3	1	1	1
Social Media, Internet Blogs and News	Ordering, Planning & CRM	Collecting data from Social Media, Internet Blogs and News and evaluating that data to make business decisions	2	2	2	2
Email Records	Ordering, Planning & CRM	Capture and registration of business email records in official recordkeeping system to ensure accountability and future decision-making process	2	2	2	2
Product Reviews	Ordering, Planning & CRM	Customers' ratings and comments on products (goods and services) they have tried, on Internet	2	2	2	1

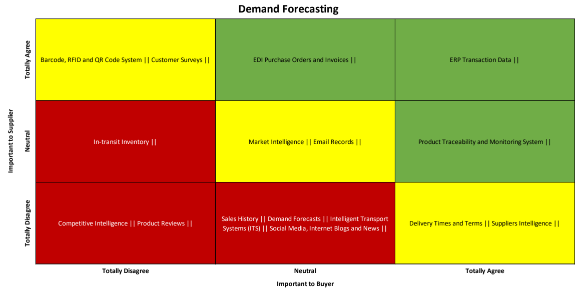
[(1)] Source: (Rozados & Tjahjono, 2014)
[(2)] Source: WikiPedia, InvestoPedia, Guru99, … etc.

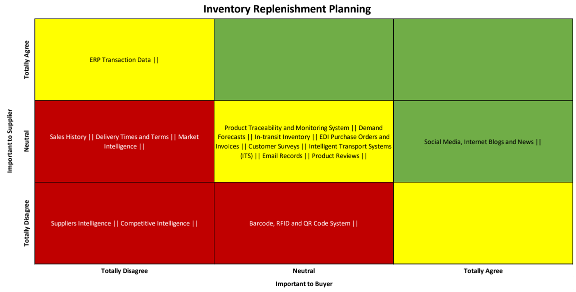
Pre-Meeting Survey - Part 1: Data Sources

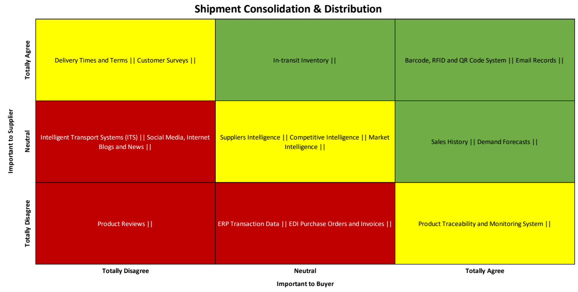
From your point of view as a **Supplier** - focusing on Order Replenishment Cycle, please indicate the perceived Importance (1 Totally Disagree - 2 Neutral - 3 Totally Agree) of sharing each of the following Data Sources between you and **your Buyer** in each of the four listed Supply Chain decision domains…

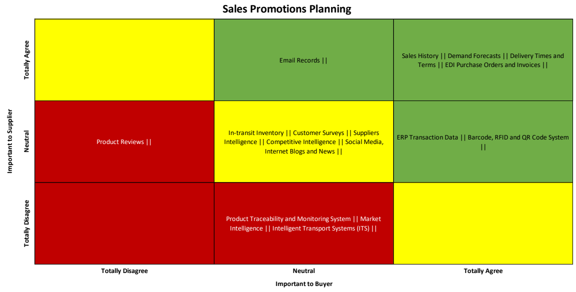
Supplier's Data Sources[(1)]	Supply Chain Function[(1)]	Data Source Description[(2)]	Perceived Important to be Shared (1 Totally Disagree - 2 Neutral - 3 Totally Agree)			
			Demand Forecasting	Inventory Replenishment Planning	Shipment Consolidation & Distribution	Sales Promotions Planning
Structured Data Sources						
ERP Transaction Data	Ordering, Planning & CRM	Transactions' information recorded on ERP (Enterprise Resource Planning)	1	3	1	2
Product Traceability and Monitoring System	Warehousing	System that follows a product through all phases of manufacturing, warehousing and distribution process	1	2	3	1
Barcode, RFID and QR Code System	Warehousing	System that automatically identify and track Barcode, RFID (Radio Frequency Identification), and QR Code (Quick Response Code) tags attached to different objects	1	2	3	1
Sales History	Ordering, Planning & CRM	Summary data of the sales of a product for a given time period (months or years)	2	2	2	2
Point-of-Sales System	Ordering, Planning & CRM	Terminals and Systems where customers execute their payment for goods or services in physical store, or virtual sales point such as a computer or mobile electronic device	2	2	1	1
Demand Forecasts	Ordering, Planning & CRM	Demand prediction of a product in the future (months or years)	2	1	2	3
Delivery Times and Terms	Procurement & SRM	Agreed timing and/or rate of delivery between a buyer and a seller, for products purchased for a future delivery period	1	2	3	1
In-transit Inventory	Warehousing	Inventory that have shipped from the seller, but not yet received by the buyer	1	1	3	3
Semi-Structured Data Sources						
EDI Purchase Orders and Invoices	Procurement & SRM	Electronic version of Invoice sent via EDI (Electronic Data Interchange), that is typically sent in response to an electronic Purchase Order including prices and quantities	2	3	3	2
Customer Surveys	Ordering, Planning & CRM	Customer answers to questions that are used to analyze whether or not changes need to be made in business operations to increase customers' satisfaction and loyalty	1	3	3	3
Suppliers Intelligence	Ordering, Planning & CRM	Ethical monitoring of strategic suppliers (e.g. suppliers' pricing, other customers, capacity, financials, … etc..)	3	2	3	2
Competitive Intelligence	Ordering, Planning & CRM	Ethical monitoring of main competitors, main customers, and any aspect of the environment needed (e.g. competitors' pricing, capacity, financials, … etc.)	2	3	3	3
Market Intelligence	Ordering, Planning & CRM	Ethical monitoring of strategic suppliers, main competitors, main customers, and any aspect of the environment needed (e.g. competitors' pricing, capacity, financials, … etc.)	2	2	1	2
Unstructured Data Sources						
Intelligent Transport Systems (ITS)	Logistics	Application of sensing, analysis, control and communications technologies to ground transportation in order to improve safety, mobility and efficiency	3	3	2	1
Social Media, Internet Blogs and News	Ordering, Planning & CRM	Collecting data from Social Media, Internet Blogs and News and evaluating that data to make business decisions	3	2	2	2
Email Records	Ordering, Planning & CRM	Capture and registration of business email records in official recordkeeping system to ensure accountability and future decision-making process	3	1	2	3
Product Reviews	Ordering, Planning & CRM	Customers' ratings and comments on products (goods and services) they have tried, on Internet	3	2	3	3

[(1)] Source: (Rozados & Tjahjono, 2014)
[(2)] Source: WikiPedia, InvestoPedia, Guru99, … etc.









Important Data Sources to both entities, and should be used as the immediate starting point of DB Sharing and BDA Usage implementation

Somehow Important Data Sources to both entities, and could be reconsidered during the DB Sharing and BDA Usage implementation planning meetings and/or in future enhancement stages

Unimportant Data Sources to both entities, and could be dropped in DB Sharing and BDA Usage implementation

Pre-Meeting Survey - Part 2: Big Data Analytics

From the **Buyer's** point of view - focusing on Order Replenishment Cycle, please indicate the perceived Importance of each of the following Big Data Analytics Methods or Techniques in each of the four listed Supply Chain decision domains…

Big Data Analytics	Big Data Analytics Description[(1)]	Important to be Used in Order Replenishment Cycle (1 Totally Disagree - 2 Neutral - 3 Totally Agree)			
		Demand Forecasting	Inventory Replenishment Planning	Shipment Consolidation & Distributic	Sales Promotions Planning
Descriptive Analytics					
Supply Chain Visualization	Visualization techniques to illustrate the relationships within data, that attempt to get the big picture, to know the unknowable, to predict the future, and to anticipate responses to the unforeseen by displaying real-time changes and using more illustrative graphics beyond traditional charts	2	3	1	1
Predictive Analytics					
Time Series Analysis	Analyzing time series data to see how a variable changes over time and how it shifts compared to other variables to extract trends, seasonality, cyclic, meaningful statistics, and other characteristics	3	1	2	3
Linear, Non-Linear, and Logistic Regression	Statistical processes to measure the relationships between a dependent variable and one or more independent variables using linear, non-linear, and/or logistic approaches	2	1	2	2
Data-Mining Techniques	Process of looking for hidden, valid, and potentially useful patterns in huge data sets involving exploring and analyzing large data blocks using methods at the intersection of machine learning, statistics, and database systems	2	2	3	2
Simulation	Representation of a problem that approximate operation of process or system over time for performance tuning or optimizing, safety engineering, testing, … etc.	2	1	2	3
Social Media and Internet Content Analysis	Form of content analysis for Internet-based communication through systematic examination and interpretation	2	3	3	2
Prescriptive Analytics					
Linear and Non-Linear Programming	Process of solving an optimization problem where some of constraints or objective function are linear or nonlinear, by calculating an objective function over a set of unknown real variables to the satisfaction of a system of equalities and inequalities, collectively termed constraints	2	2	3	2
Mixed-Integer Linear Programming (MILP)	Process of solving an optimization problem where all constraints and objective function are linear, by calculating an objective function over a set of unknown integer variables to the satisfaction of a system of equalities and inequalities, collectively termed constraints	2	2	1	2
Network Flow Algorithms	Process of solving an optimization problem based on graph theory, by modeling anything travels through a network of nodes that receives or sends a flow that must satisfy restrictions or capacity constraints	2	1	1	3
Stochastic Dynamic Programming	A technique for modelling and solving problems under uncertainty, that aims to compute how to act optimally in the face of uncertainty and to maximize expected reward over a planning horizon, and it deals with problems in which the current period reward and/or the next period state are random	1	1	2	2
Other Analytics (Please Specify)					

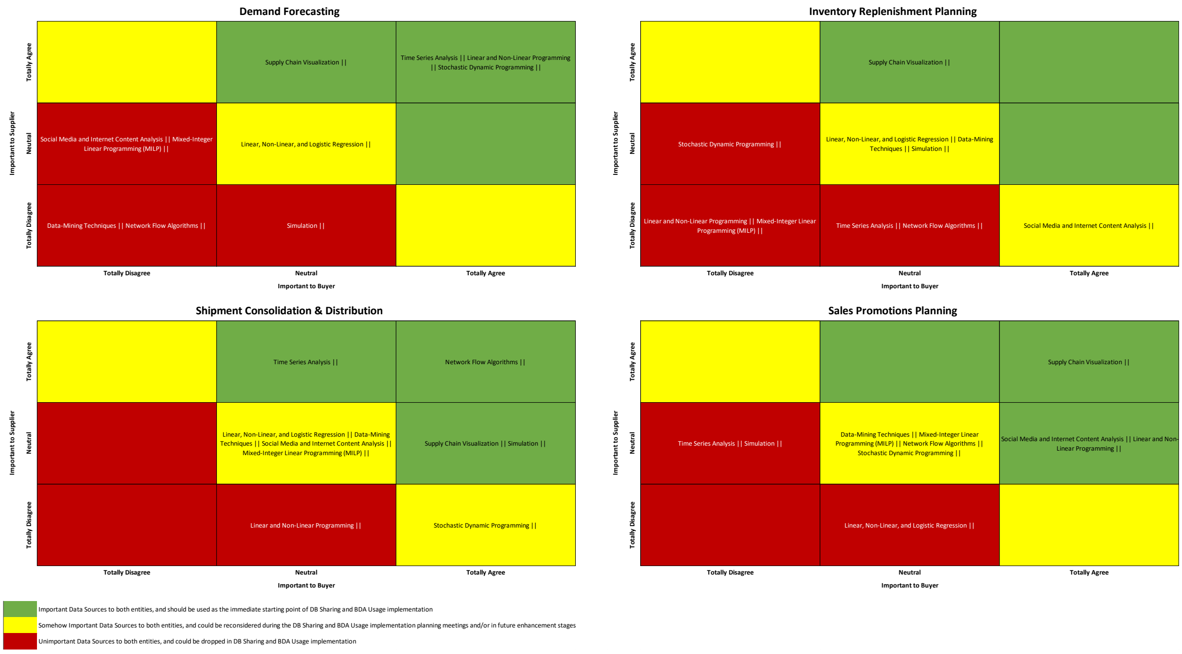
[(1)] Source: WikiPedia, InvestoPedia, Guru99, … etc.

Pre-Meeting Survey - Part 2: Big Data Analytics

From the **Supplier's** point of view - focusing on Order Replenishment Cycle, please indicate the perceived Importance of each of the following Big Data Analytics Methods or Techniques in each of the four listed Supply Chain decision domains…

Big Data Analytics	Big Data Analytics Description[(1)]	Important to be Used in Order Replenishment Cycle (1 Totally Disagree - 2 Neutral - 3 Totally Agree)			
		Demand Forecasting	Inventory Replenishment Planning	Shipment Consolidation & Distribution	Sales Promotions Planning
Descriptive Analytics					
Supply Chain Visualization	Visualization techniques to illustrate the relationships within data, that attempt to get the big picture, to know the unknowable, to predict the future, and to anticipate responses to the unforeseen by displaying real-time changes and using more illustrative graphics beyond traditional charts	3	2	2	1
Predictive Analytics					
Time Series Analysis	Analyzing time series data to see how a variable changes over time and how it shifts compared to other variables to extract trends, seasonality, cyclic, meaningful statistics, and other characteristics	3	2	1	2
Linear, Non-Linear, and Logistic Regression	Statistical processes to measure the relationships between a dependent variable and one or more independent variables using linear, non-linear, and/or logistic approaches	1	2	2	2
Data-Mining Techniques	Process of looking for hidden, valid, and potentially useful patterns in huge data sets involving exploring and analyzing large data blocks using methods at the intersection of machine learning, statistics, and database systems	3	2	3	2
Simulation	Representation of a problem that approximate operation of process or system over time for performance tuning or optimizing, safety engineering, testing, … etc.	2	1	3	2
Social Media and Internet Content Analysis	Form of content analysis for Internet-based communication through systematic examination and interpretation	3	2	3	3
Prescriptive Analytics					
Linear and Non-Linear Programming	Process of solving an optimization problem where some of constraints or objective function are linear or nonlinear, by calculating an objective function over a set of unknown real variables to the satisfaction of a system of equalities and inequalities, collectively termed constraints	3	2	1	2
Mixed-Integer Linear Programming (MILP)	Process of solving an optimization problem where all constraints and objective function are linear, by calculating an objective function over a set of unknown integer variables to the satisfaction of a system of equalities and inequalities, collectively termed constraints	2	3	3	2
Network Flow Algorithms	Process of solving an optimization problem based on graph theory, by modeling anything travels through a network of nodes that receives or sends a flow that must satisfy restrictions or capacity constraints	3	1	3	2
Stochastic Dynamic Programming	A technique for modelling and solving problems under uncertainty, that aims to compute how to act optimally in the face of uncertainty and to maximize expected reward over a planning horizon, and it deals with problems in which the current period reward and/or the next period state are random	2	2	1	2
Other Analytics (Please Specify)					

[(1)] Source: WikiPedia, InvestoPedia, Guru99, … etc.


Demand Forecasting
Important to Supplier
Totally Agree
Neutral
Totally Disagree
Supply Chain Visualization ||
Time Series Analysis || Linear and Non-Linear Programming || Stochastic Dynamic Programming ||
Social Media and Internet Content Analysis || Mixed-Integer Linear Programming (MILP) ||
Linear, Non-Linear, and Logistic Regression ||
Data-Mining Techniques || Network Flow Algorithms ||
Simulation ||
Totally Disagree
Neutral
Totally Agree
Important to Buyer
Inventory Replenishment Planning
Important to Supplier
Totally Agree
Neutral
Totally Disagree
Supply Chain Visualization ||
Stochastic Dynamic Programming ||
Linear, Non-Linear, and Logistic Regression || Data-Mining Techniques || Simulation ||
Linear and Non-Linear Programming || Mixed-Integer Linear Programming (MILP) ||
Time Series Analysis || Network Flow Algorithms ||
Social Media and Internet Content Analysis ||
Totally Disagree
Neutral
Totally Agree
Important to Buyer
Shipment Consolidation & Distribution
Important to Supplier
Totally Agree
Neutral
Totally Disagree
Time Series Analysis ||
Network Flow Algorithms ||
Linear, Non-Linear, and Logistic Regression || Data-Mining Techniques || Social Media and Internet Content Analysis || Mixed-Integer Linear Programming (MILP) ||
Supply Chain Visualization || Simulation ||
Linear and Non-Linear Programming ||
Stochastic Dynamic Programming ||
Totally Disagree
Neutral
Totally Agree
Important to Buyer
Sales Promotions Planning
Important to Supplier
Totally Agree
Neutral
Totally Disagree
Supply Chain Visualization ||
Time Series Analysis || Simulation ||
Data-Mining Techniques || Mixed-Integer Linear Programming (MILP) || Network Flow Algorithms || Stochastic Dynamic Programming ||
Social Media and Internet Content Analysis || Linear and Non-Linear Programming ||
Linear, Non-Linear, and Logistic Regression ||
Totally Disagree
Neutral
Totally Agree
Important to Buyer
Important Data Sources to both entities, and should be used as the immediate starting point of DB Sharing and BDA Usage implementation
Somehow Important Data Sources to both entities, and could be reconsidered during the DB Sharing and BDA Usage implementation planning meetings and/or in future enhancement stages
Unimportant Data Sources to both entities, and could be dropped in DB Sharing and BDA Usage implementation

DECISION SCIENCES INSTITUTE

Factors affecting customer satisfaction in online video streaming

Shameem Shagirbasha
Great Lakes institute of Management
Email: shawaja@gmail.com

ABSTRACT

In this research we have developed and empirically tested a model for explaining/predicting customer satisfaction in online video streaming services. One unique feature about our study is the role peer influence plays in subscription of streaming services. More and more people purchase these subscriptions after hearing positive reviews about particular TV shows or movies specific to one subscription platform from their friends/ acquaintances. Form this study we will identify if there is any significant relation between satisfaction and peer influence. The empirical results show that peer influence and originality of content are significant variables that affect customer satisfaction in subscription-based video streaming services and content quality was found to be marginally significant.

Keywords: customer satisfaction, online video streaming, peer influence

INTRODUCTION

Due to the development and mutual convergence of broadcasting and communication technologies, the media industry has faced a completely different environment from the past. In the age of multi-media diversity, media users who have received information unilaterally and participated only passively are becoming more and more. It is changing the behaviour of the users who have relied only on the real-time viewing through the TV receiver to select and view the desired content at the desired time through the general network of the Internet instead of the existing broadcasting network using various digital devices.
Traditional TV channel users have only one-sided choice because of the limitations of time and channel consumption. As the choice is limited, there is a clear limit to the qualitative growth of broadcasting services. In addition, the importance of smart phones in everyday life has become increasingly important, and smartphones have been recognized as a necessary medium for media services.

Entertainment Industry recently is witnessing a change in dynamics with the entrance of OTT (Over the top) segment. In the age of digitalization, customers watch entertainment content- Movies, TV shows, music directly over the internet as a standalone product instead of television through cable networks and broadcast channels. In this process, OTT service is the most suitable service for media usage. The OTT service is defined as a service that provides a video service to a digital device by using a general- purpose Internet network. OTT service, which is the research target of this study, has characteristics of personalized bi-directional media service rather than traditional unidirectional media. Since the OTT service is a service using the general internet network, it is possible to use the contents desired by the user anywhere on the Internet at any time (any time, any device, anywhere). These characteristics have had a great impact on the spread of personal broadcasting and online video services such as UCC (User Create Contents) and video clips

The global OTT devices and services market is expected to reach USD 165.13 billion by 2025, according to the research by Research and Market [1]. Moreover, this research also suggests that the service providers of OTT across the globe are focusing on developing the three key streaming technologies viz. Virtual reality video, 4K streaming, and High dynamic range. This highly competitive market and high demand will result in tremendous growth of this industry. OTT devices' technologies would also be growing at a high rate. Also, the growth of smart TVs is expected to bring in enormous opportunities for advertisers which would compete to become the content providers, offering a wide range of services to the customers through various platforms like Netflix, YouTube, Hulu, etc. [1] As far as the India is concerned, OTT market is expected to grow at a CAGR of 23%. For the year 2017, India stood 3rd in the list of top 10 countries for video on demand services by revenue. India is expected to be the 10th largest market in terms of revenue by 2020 as the revenues are expected to double from 406 mn USD in 2018 to 805mn USD in 2022. (PwC global industry and media outlook).

LITERATURE REVIEW

Customer Satisfaction:

Customer satisfaction is a term used to quantify whether a product or service meet or exceed the customer demand. Initial studies on customer satisfaction focussed on Expectancy disconfirmation theory, which states that satisfaction is a function of expectation, perceived performance and disconfirmation of beliefs. If a product or service is better than expected, it leads to a positive disconfirmation which leads to satisfaction. In other words, Positive disconfirmation leads to satisfaction and negative disconfirmation leads to dissatisfaction (Oliver and DeSarbo, 1988)

A study on customer satisfaction with e services (Mohamed Khalifa, 2002) included a desire disconfirmation factor in their model. Similar to previous research (Oliver and DeSarbo, 1988), positive disconfirmation of desire leads to customer satisfaction. All the variables taken under the desire factor- Information quality, membership service quality and user friendliness were found to be statistically significant for improving customer satisfaction. As a part of our study, we intend to include desire factors for determining customer satisfaction

Service Quality:

Service quality plays a vital role in the success of an organization. Customer satisfaction and their trust are vital in terms of customer retention when it comes to services industry. This could only be achieved by delivering good quality service to the customers and providing customer perceived service quality. Results of (Saba M. Hussain Int. Journal of Engineering Research and Applications, March 2014) indicates that in case of e services Reliability, Responsiveness, Ease of Use, Personalization and Website Design, each has significant influence on the Customer Perceived service quality Service quality is an indirect determinant of customer satisfaction. Customer's perception of electronic service quality is different from the traditional approach (Parsuraman et al,).

Zeithaml, Parasuraman, and Malhotra's in 2000 identified attributes to measure the service quality. Similar constructs were identified by Mohamad Ibrahim Ladan, 1008 for Web services. Below are the variables that can be relevant for online video streaming services and their explanation:

1. Performance: Performance of website is a significant quality metric. The high page load speed, low latency of data transfer, low response time for transaction enhance

service quality. High throughput, or the number of interactions handled by a web page per second is also a performance metric which can be used for OTT service since they are web based and the performance metrics can be used to measure buffer speed, browsing speed and response time of website after an action

2. Accessibility: It measures the ability to use web service whenever needed. The OTT service should be easily accessible on multiple devices. The service provider should have mobile and desktop applications meeting the expectations. Even a separate mobile website is provided by company.
3. Customization Personalization: The website should be able to transform itself on the basis of individual user's preferences (A. Parasuraman). OTT providers like Netflix and Amazon prime provide personalized screen on the basis of user's history. While user creates their account, Netflix asks certain set of questions and on its basis prepare a personalized screen of movies and tv series that the person would like. Personalization can be considered as a desire factor of customer satisfaction study as it leads to user friendliness

H1: There is a positive relationship between Performance of the website and Customer satisfaction

H2: There is a positive relationship between Accessibility/Ease of Use of the website and Customer satisfaction

H3: There is a positive relationship between effective personalization of the website and Customer satisfaction

SERVQUAL dimensions (Parsuraman et al, 1988) determine Core product or service as an important measure of quality. Further studies (G.S. Sureshchandar, Chandrasekharan Rajendran, R.N. Anantharaman, 2002) also found Core product quality to be a statistically significant parameter in determining customer satisfaction.

Content is the core offering of video service providers and according to a research (Harry Bouwman, Abo Akademi University, Turku, Finland, and Delft University of Technology,2015), News and Sports correlates with customer satisfaction. Customers who are sports enthusiast or like being updated with the happenings around him/her, need these services to have anytime access. Video service platform such as Hotstar is used by such customers for watching sports or news. This seems to be a significant factor that drives customer satisfaction according to the paper (Over the top mobile TV: A field trial by public and commercial broadcasters). The content strategy of Netflix consists of three parameters, Originality of content, Popular TV shows and Movies with higher ratings (Mandal, Diroma, Jain ,2017) and these are key drivers for their growth which eventually comes from customer satisfaction and retention

H4: There is a positive relationship between Originality of Content offered and satisfaction of customers of the video service provider

H5: There is a positive relationship between perceived quality of content and customer satisfaction

Price

Although the price dimension is not specified in ES-QUAL dimension, Johan Ström, Kelly Bueno Martínez,2013 concluded that price is an important factor determining perceived quality and eventually satisfaction. Satisfaction can also be a measure of willingness to pay of the customer. India being a price sensitive market, we will take perceived price variable under consideration. No customer wants to pay more and get a service which they could easily get for a lower price. Hence, we have found that a customer is more satisfied with an offer of good service with a lower price. This leads to our fourth hypothesis.

H6: There is a relationship between the price that the customer is charged and his/her satisfaction

Video Quality

Video based services are moving towards providing 4K and UHD content to their subscribers. In a span of 4 to 7 years, UHD services will become conventional in media sector. Key drivers for upgrade to high quality video will be movies and sports content. OTT segment will be second in the list to gain momentum in high quality video. (Global Survey for Media Executives).
OTT services offer different prices for different video qualities. UHD video quality is available in its premium subscription plans. High Quality videos in 4K and UHD are priority investments areas for most OTT service providers to improve customer satisfaction (Akamai technologies,2016). Availability of high definition programs drive users to pay for subscription based streaming services (Accenture Digital Consumer Survey, 2018). From the above findings we can conclude that video quality is a significant factor that customers consider while purchasing a plan in subscription-based video streaming services

H7: There is a positive relationship between video quality and customer satisfaction

Peer Influence
Trust is crucial when it comes to services. It has been defined differently in different context and have been extensively investigated. It could be, in very simple terms, defined as the relying upon someone. E Services involve a trust element in them. Gaining a Customer's trust is another way of gaining new customers. Only trustworthy customers would be loyal and hence become the Promoters of the service provider. The NPS is a vital metric for the growth of an organization. Hence, the same applies to the video service providers too. The existing customers' feedback should be taken and more focus should made on converting the passively satisfied customers to Promoters. According to (Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1489–1498), an individual's behaviour could be extensively affected by Peer Influence. The basic tendency of a human is to conform to group he/she is around. Peer Influence could change the attitude and perception of individuals towards services/products. Hence, we could relate the same with the video streaming service providers, more the promoters, more the number of new customers.
Peer influence extensively changes the attitude of an individual and from our studies we have found that people tend to conform to the groups they belong to. Hence, we propose that is there is a good correlation between Peer Influence and Customer Satisfaction.

H8: There is a positive relationship between Peer influence and Customer Satisfaction

METHOD

We have conducted a survey with questions on these eight variables, with three questions for each construct to ensure reliability of data. We would then proceed with applying various techniques to clean the data and run a regression on SPSS or relevant tool to conclude. Every aspect related to the website services like quality of content, performance of website, video quality, accessibility, price, peer influence, and originality of content are studied at different level of the research. We conducted a Focused Group Discussion in order to get better insights from the people. The FGD method and literature reviews were used to prepare a questioner. A group of three students contacted more than 20 existing users, of which 3 groups were formed to collect the relevant data. Participants were having heterogeneity among various parameters like gender and age.

A structured questionnaire was designed using a 5-point Likert scale after collecting the data from the participants and the literature reviews. The questionnaire covers all the independent variables and the depend variable i.e. customer satisfaction. The questionnaire was floated to about 30 users to do a pilot survey. This needs to be done to check whether the questioners are relevant and are we going on the right track. After the pilot survey, we circulated the questionnaire to a population consisting of existing users of online video streaming websites. SPSS would be used to find the Cronbach's alpha – a measure of relevance of the data. It is used to check the reliability and relevance of each construct by ensuring the Cronbach's Alpha was above the minimum acceptable level of 0.6.
This will run for all the items under each construct. We introduced relevant covariates like age and gender to evaluate how this affects out results. After this, we will find the correlation between each variable. If there is any significant correlation found, then factor analysis will be used to eliminate the correlation between the variables. Then we will run a regression and interpret the results to find significant factors.

RESULT

The data for different item was cleaned. The Cronbach alpha for all the constructs was calculated and was found to be within acceptable limits for 7 of the 8 variables. For one variable, peer influence, we deleted one item to get the Cronbach alpha value to .671
Correlations between different values was then performed. The Pearson correlation coefficient was found to be less than 0.6 and hence we did not perform factor analysis
We performed linear regression for which adjusted R square value was found to be 0.567
Peer influence and Originality of content were found to be significant variables that influence customer satisfaction for subscription-based video streaming services. The beta value were also high for these variables. Content quality was marginally significant with high beta value. Gender did not play a significant role in customer satisfaction

Reliability Analysis

DV : Satisfaction

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.852	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Satisfaction Q1	8.04	2.364	.721	.798
Satisfaction Q2	8.23	2.286	.764	.760
Satisfaction Q3	7.97	1.901	.708	.826

<u>IV1: Accessibility</u>

Reliability

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.886	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Accessibility Q1	6.78	4.674	.771	.845
Accessibility Q2	7.09	4.385	.789	.828
Accessibility Q3	6.84	4.153	.777	.841

<u>IV2: Price</u>

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.865	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Price Q1	6.01	4.705	.743	.813
Price Q2	6.09	4.242	.772	.785
Price Q3	6.13	4.616	.719	.833

IV3: Performance

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.712	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Performance Q1	8.24	2.451	.540	.614
Performance Q2	8.52	1.984	.570	.578
Performance Q3	8.07	2.584	.496	.665

IV4: Video Quality

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.762	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Video Quality Q1	7.64	3.305	.482	.793
video Quality Q2	8.16	2.296	.715	.528
Video Quality Q3	8.22	2.674	.600	.672

IV5: Content Quality

Reliability

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.839	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Content Quality Q1	7.48	3.698	.692	.786
Content Quality Q2	7.58	3.604	.740	.740
Content Quality Q3	7.64	3.626	.676	.803

IV6: Originality

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.852	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Originality Q1	7.88	3.431	.641	.871
Originality Q2	7.98	3.125	.774	.743
Originality Q3	7.91	3.385	.760	.761

IV7: Peer influence

a. Listwise deletion based on all variables in the procedure.

Case Processing Summary

		N	%
Cases	Valid	113	100.0
	Excluded[a]	0	.0
	Total	113	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha[a]	N of Items
-.186	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Peer Influence Q1	5.7788	1.495	.104	-.764[a]
Peer influence Q2	6.0177	.803	.267	-2.175[a]
PeerInfluence_Q3_transformed	7.7257	3.326	-.452	.671

a. The value is negative due to a negative average covariance among items. This violates reliability model assumptions. You may want to check item codings.

IV8: Personalization

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.582	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Personalization Q1	8.27	2.147	.357	.536
Personalization Q2	7.68	2.451	.300	.610
Personalization Q3	8.03	2.008	.533	.265

Correlations

	Accessibility	Price	Performance	Video Quality	Content Quality	Originality	Peer influence	Personalization
Accessibility	0.886	.741**	.339**	.341**	.268**	.344**	.281**	.312**
Price	.741**	0.865	.341**	.285**	.249**	.251**	.256**	.257**
Performance	.339**	.341**	0.712	.644**	.429**	.323**	.411**	.396**
Video Quality	.341**	.285**	.644**	0.762	.461**	.357**	.431**	.524**
Content Quality	.268**	.249**	.429**	.461**	0.839	.683**	.549**	.553**
Originality	.344**	.251**	.323**	.357**	.683**	0.852	.499**	.561**
Peer influence	.281**	.256**	.411**	.431**	.549**	.499**	0.671	.469**
Personalization	.312**	.257**	.396**	.524**	.553**	.561**	.469**	0.582

**. Correlation is significant at the 0.01 level (2-tailed).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.776[a]	.602	.567	.4692026

a. Predictors: (Constant), Gender Cont, Personalization, Price, Performance, Peer influence, Originality, Video Quality, Accessibility, Content Quality

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.243	9	3.805	17.282	.000[b]
	Residual	22.676	103	.220		
	Total	56.918	112			

a. Dependent Variable: Satisfaction Y
b. Predictors: (Constant), Gender Cont, Personalization, Price, Performance, Peer influence, Originality, Video Quality, Accessibility, Content Quality

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.784	.321		2.443	.016
	Accessibility	.109	.068	.156	1.608	.111
	Price	-.038	.065	-.055	-.585	.560
	Performance	.077	.085	.078	.914	.363
	Video Quality	.076	.082	.084	.936	.352
	Content Quality	.125	.076	.161	1.645	.103
	Originality	.161	.075	.198	2.135	.035
	Peer influence	.271	.062	.347	4.361	.000
	Personalization	.067	.090	.064	.747	.457
	Gender Cont	-.147	.095	-.101	-1.537	.127

a. Dependent Variable: Satisfaction Y

DISCUSSION

The above empirical study was conducted to determine significant variables in online video streaming services. After completion of literature review and qualitative research, we came across the following list of variables which drive customer satisfaction:

1. Accessibility: The ease of use of service anywhere, anytime
2. Price: Refers to the affordability and value for money for the consumers
3. Performance: A good interface that provides the consumer a great experience while

consuming the service

4. Video Quality: The appealing quality of the videos
5. Content Quality: Including variety of the top-rated movies and shows
6. Originality of content: Exclusive offerings for the consumer
7. Peer Influence: How friends/family could drive someone's choice of the service
8. Personalization: How significant is personalization provided by the service provider in

driving customer satisfaction

The results indicate that peer influence and originality of content are significant variables in determining customer satisfaction in online video streaming services. Another variable i.e. content quality was also found to be marginally significant. Today the number of streaming service providers have increased and the market is hence becoming a red ocean. In this competitive market, survival is going to be difficult if the service is not customer centric. Consumers look for an experience and not just a mere service when they go to a service provider. So, a service provider with a blend of value plus a good experience for its consumer could survive and grow well.

REFERENCES

Over-the-Top (OTT) Devices and Services Market Analysis by Content (VoIP, Text & Images, Video), by Revenue Source (AVOD, SVOD, TVOD), by Device, by Deployment, And Segment Forecasts, 2014 – 2025. Research and Markets

Richard L. Oliver and Wayne S. DeSarbo, Response Determinants in Satisfaction Judgments, Source: The Journal of Consumer Research, Vol. 14, No. 4 (Mar., 1988), pp. 495-507

Mohamed Khalifa, Vanessa Liu “Satisfaction with Internet-Based Services”,2002
Zeithaml, Parasuraman, and Malhotra “E-S-QUAL A Multiple-Item Scale for Assessing Electronic Service Quality” , 2000

A Conceptual Model of Service Quality and Its Implications for Future Research Author(s): A. Parasuraman, Valarie A. Zeithaml, Leonard L. Berry Source: The Journal of Marketing, Vol. 49, No. 4 (Autumn, 1985), pp. 41-50 Published by: American Marketing Association

Saba M. Hussain Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 3(Version 1), March 2014, pp.350-359
OVER THE TOP MOBILE TV: A FIELD TRIAL BY PUBLIC AND COMMERCIAL BROADCASTERS, Harry Bouwman, Åbo Akademi University, Turku, Finland, and Delft (University of Technology) Mark de Reuver, Delft University of Technology, Schalk Stalman, Triple IT

Factors affecting the intention to use e-Government services (Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1489–1498)
“The determinants of customer satisfaction, loyalty and willingness to pay in subscription based streaming services”, Department of Marketing and Strategy, Stockholm School of Economics, 2013

G.S. Sureshchandar, Chandrasekharan Rajendran, R.N. Anantharaman, (2002) "The relationship between service quality and customer satisfaction – a factor specific approach", Journal of Services Marketing, Vol. 16 Issue: 4, pp.363-379

Md. Ariful Islam, Proshenjit Ghosh, Md. Rayhan Islam, Mithun Sarker. Customers' Satisfaction about Services Quality & Services
Provided by BASIC Bank Ltd. - A Study on Khulna. International Journal of Business and Management. Vol. 1, No. 4, 2013, pp. 58-73. doi: 10.11648/j.sjbm.20130104.12

Mandal, G., Diroma, F., Jain, R. (2017). Netflix: An In-Depth Study of their Proactive & Adaptive Strategies to Drive Growth and Deal with Issues of Net-Neutrality & Digital Equity. IRA-International Journal of Management & Social Sciences (ISSN 2455-2267), 8(2), 152-161. doi:http://dx.doi.org/10.21013/jmss.v8.n2.p3

Grace T.R. Lin, Chia-Chi Sun, (2009) "Factors influencing satisfaction and loyalty in online shopping: an integrated model", Online Information Review, Vol. 33 Issue: 3, pp.458-475

Jordan, Andrew K., "The Effects of Netflix and Blockbuster Strategies on Firm Value" (2011). CMC Senior Theses. Paper 154. http://scholarship.claremont.edu/cmc_theses/154

Mohamad Ibrahim Ladan “Web Services Metrics: A Survey and A Classification” 2011 International Conference on Network and Electronics Engineering IPCSIT vol.11 (2011)

Accenture Digital Consumer Survey, 2018

Akamai technologies, Measuring Video Quality and Its Impact to Your Bottom Line A Customer Innovation Story with WUAKI.TV, a Rakuten company 2016

Zemblytė, J. (2015). The instrument for evaluating e-service quality. *Procedia-Social and Behavioral Sciences*, *213*, 801-806.

DECISION SCIENCES INSTITUTE

Predicting Adoptions and Sales of Technology Products: A Fractional Calculus-Based Approach

Aslan Lotfi
University of Richmond
Email: alotfi@richmond.edu

Zhengrui Jiang
Nanjing University
Email: zjiang@nju.edu

Ali Lotfi
Western University
Email: alotfi2@uwo.ca

Dipak Jain
China Europe International Business School
Email: dipakcjain@ceibs.edu

ABSTRACT

Classic product diffusion models consider only initial product purchases, therefore making them ill-fitted for sales prediction for today's technology products with frequent repeat purchases. Despite the long tradition of product diffusion research, there exists no viable model option when repeat purchases constitute a large proportion of product sales. The present study introduces a new sales growth model to fill this void. Empirical results show that the new model fits sales data with varying proportions of repeat purchases. In addition, the proposed model can be extended to incorporate marketing mix variables, further enhancing its applicability in the prediction of product sales.

KEYWORDS: Diffusion of innovation, Repeat purchases, Replacements, Multi-unit ownerships, and Fractional calculus

INTRODUCTION

Predictive analytics are playing an important role in organizations' decision-making processes, thus increasingly embraced by the information systems community (e.g., Shmueli & Koppius, 2011). In particular, accurately predicting the sales trajectory of a product in its lifecycle is critically important for firms' operational and strategic planning such as deciding the medium- and long-term production, distribution, and after-sale service capacities, as well as acquiring and allocating financial resources to support critical business operations.

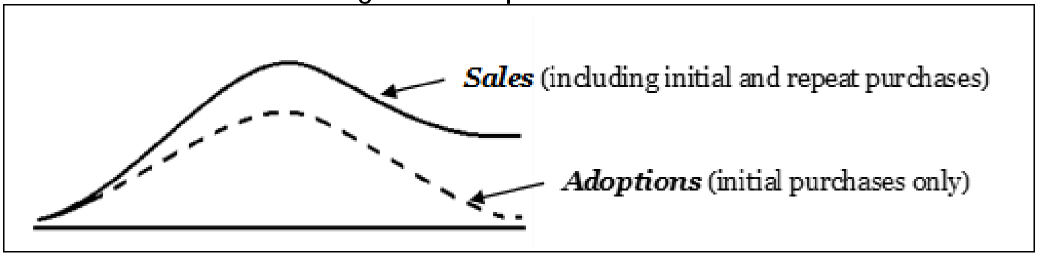
Since empirical evidence has shown that most products' sales growth rate in a lifecycle follows a bell-shaped curve, or equivalently, the cumulative sales follows an S-shaped curve (Rogers, 2003), numerous bell- or S-shaped diffusion models have been proposed to analyze and predict the sales trajectory in a product lifecycle (Mahajan et al, 2000; Bass 2004). The most well-known one among the known diffusion models is the Bass Model (Bass, 1969). Since its inception, thanks to its great empirical performance and ease of implementation, the Bass Model has been applied by researchers from various disciplines such as marketing, operations, and information systems (IS). In the IS literature, for instance, the Bass Model and its extensions have been applied to model the growth of information systems spending (Gurbaxani & Mendelson, 1990), information systems outsourcing (Loh & Venkatraman, 1992; Hu et al, 1997), the sales of mainframe computing (Tam, 1996), mobile telecommunication services (Niculescu & Whang, 2012), multigeneration software platforms (Hann et al, 2016), free software products (Jiang & Sarkar, 2009), and multigeneration technology products (Guo & Chen, 2018; Jiang et al, 2019).

It is important to note that the diffusion models, represented by the Bass Model, count only first-time adoptions of a product. For instance, when a consumer purchases a product for the first time, the consumer turns irreversibly from a “non-adopter” to an “adopter.” Subsequently, if the consumer repurchases the same hardware product or renews subscription to a software product in the future, the repeat purchase(s) will not be recounted as new adoptions. In fact, to avoid any confusion, Bass clarifies in his seminal work that the Bass Model is concerned “only with the timing of the initial purchase.” (Bass, 1969, p. 215). For this reason, we use adoptions, initial purchases, and first-time purchases interchangeably in this study.

Since diffusion models capture first-time purchases only, they can be used to estimate sales only when repeat purchases are insignificant. Therefore, Bass (1969) and many subsequent studies test the Bass Model using sales data for consumer durables such as TV sets and washing machines, of which repeat purchases are infrequent at least in the initial time periods, and for the diffusion of technologies/practices (e.g., agricultural technologies, outsourcing) for which recounting generally does not occur (Mahajan et al, 2000; Hu et al, 1997). Today’s technology products, including both hardware and software products, however, often have frequent releases of (sometimes marginally improved) new versions/generations (e.g., 1-2 years for iPhones or many famous video games), and some consumers tend to keep buying or renewing subscriptions to new product generations. Furthermore, unlike decades ago, consumers owning multiple units of the same (or similar) technology products is becoming increasingly common (e.g., multiple laptop or tablet computers in a household, multiple versions of a video game). In the presence of such frequent repeat purchases, traditional diffusion models are no longer applicable for use in predicting a product’s sales trajectory in its lifecycle. Therefore, the current era of fast-paced technology advancement and increasing consumer purchasing power calls for a sales growth model that counts both initial purchases as well as repeat purchases.

We next illustrate how sales growth rates differ with and without repeat purchases. If repeat purchases are insignificant, or we count only initial purchases (or adoptions), the market saturation effect comes into play; hence, after a market peak is reached, the rate of adoptions drops monotonically and asymptotically approaches zero. This trend is illustrated by the symmetric dashed curve shown in Figure 1. If repeat purchases are significant and counted, the growth of sales often exhibits an asymmetrical bell-shaped pattern, as illustrated by the solid curve shown in Figure 1. Even though the Bass Model does not consider repeat purchases, Bass acknowledges the existence of an asymmetric sales curve when replacement purchases are counted (see Figure 1 in Bass 1969, p. 215), and provides an accurate description of such a curve: “sales often grow to a peak and then level off at some magnitude lower than the peak” and the “stabilizing effect is accounted for by the relative growth of the replacement purchasing component of sales and the decline of the initial purchase component” (Bass, 1969, p. 215).

Figure 1. Adoptions vs. Sales



Even though Bass acknowledges that his model is designed for initial purchases only, the Bass Model is frequently applied to fit sales data that counts both initial purchases and repeat purchases. When significant repeat purchases exist, the model fitting and prediction accuracy of the Bass Model may be poor, particularly when the sales trend begins to assume an asymmetrical shape after the peak point of diffusion. Even if the model fits and predicts well in some situations despite the existence of considerable repeat purchases, the parameter estimates would be biased or misleading.

Given the demand, it seems intuitive that researchers would have developed robust sales growth models to account for frequent repeat purchases. Surprisingly, after an exhaustive literature search, we find that this is hardly the case. As we will document in the next section, a limited number of attempts have been made to model repeat purchases, and the resulting models all have some clear limitations. Based on our own investigation, we conclude that the lack of robust repeat purchase models can be attributed to the challenges in developing a repeat purchase component of sales that varies with time and can be seamlessly incorporated with the adoption component of sales. Given the clear void in the literature, the primary objective of this study is to develop a robust sales model that incorporates both adoptions and repeat purchases for a wide variety of technology products.

The rest of this paper is organized as follows. In the next section, we review the related literature. Next, we define repeat purchases. After that, we present the detailed development of the GDMR followed by a section in which we conduct the empirical testing. Conclusions and related discussions are summarized in the last section.

PRIOR LITERATURE ON DIFFUSION MODELS WITH REPEAT PURCHASES

Because of the vast body of literature on the diffusion of innovations and new product diffusions (Rogers, 2003; Mahajan et al, 2000), we focus our literature review on extended diffusion models with repeat purchases. To start, researchers have proposed model extensions to account for repeat product purchases made to replace existing product units or for adopting multiple product units. Olson and Choi (1985) propose a model assuming that sales comprise only adoptions and replacements, and that the replacement hazard function follows the Rayleigh distribution. The Olson-Choi Model is developed for cases in which the number of products in use is available.

Kamakura and Balasubramanian (1987) propose a similar model that generates long-term forecasts by incorporating the adoption and replacement components of sales. Benefiting from a more flexible hazard function to capture replacement purchases, their model is applicable with or without data for replacement sales. Their model uses information from similar products when data for replacement purchases is not available. Steffens and Balasubramanian (1998) further advance the modeling of replacement sales by allowing the distribution of the service life of replaced products to vary over time.

Focusing on the PC processor industry, Gordon (2009) introduces a dynamic structural model that explicitly considers product replacement decisions under uncertain future product quality and price. Gordon's model concentrates specifically on product replacements due to product upgrade releases. The model requires highly specific data to function as it uses a composite dataset including sales, ownership, price, and product quality. Due to data limitations, Gordon does not consider multi-unit ownership purchases (e.g., one household owning multiple laptop or tablet computers).

Lee and Hun (2017) introduce a replacement diffusion model for the growth dynamics of technologies and evaluate the performance of their model using sales data of mobile handsets in South Korea. Lee and Hun's model does not include multi-unit ownerships in their analysis.

The aforementioned models have their focus on adoptions and replacement purchases. There are other models from the prior literature that have considered other components of sales, in particular multi-unit ownership purchases. Dodson and Muller (1978) propose a model that captures the type of asymmetric sales trend illustrated by the solid curve in Figure 1 without decomposing sales into multiple components. The market they consider is composed of three groups, i.e., those who are not aware of the product, those who are aware of the product but have not made a purchase, and those who have purchased the product. Although the portrayed interaction between adopters and non-adopters is insightful, their model cannot be operationalized if the data for the different market groups is not available.

Based on econometric and simulation models, Bayus et al. (1989) incorporate first time sales, replacement sales, additional-unit sales, and institutional sales into their analysis of color television set sales. As stated by Steffens (2003), Bayus et al.'s model is developed to perform well over short terms. Steffens (2003) presents a model for sales resulting from multi-unit ownership purchases based on the Bass Model. His model differs from Bayus et al.'s in that it imposes a saturation level on multi-unit ownerships, which makes the model applicable for longer time frames.

The abovementioned studies focus primarily on repeat purchases of durable products. Repeat purchases of other product categories have also been studied in the literature. For example, Bhagat et al. (2018) model repeat purchase recommendations of consumable products. They suggest the lognormal distribution for capturing time intervals between repeat purchases. Their model, however, does not consider the diffusion of products, hence it differs significantly from those built on diffusion models.

Researchers have also developed models for repeat purchases of products for a specific industry. For instance, Lilien et al. (1981) propose a highly specialized model to project sales of prescription drugs as a function of a focal pharmaceutical company's own detailing effect, competitors' detailing effect, and word of mouth. In a subsequent study, Rao and Yamada (1988) provide further empirical support for the model by Lilien et al., then propose an alternative method for developing priors for the new drug's parameters, and demonstrate how the parameters can be updated after sales data becomes available. Because these two existing models are specifically designed for prescription drugs, they cannot be used to predict sales of other product categories.

Our review of the literature suggests that there is a need for a comprehensive model that can fit aggregate sales data that records both initial purchases and repeat purchases including both replacements and multi-unit ownerships. The models presented in the extant literature are applicable mostly when separate data for distinct sales components are available; however, in practice, most of the times only aggregate product sales data is available. Therefore, in this study, we develop a model that can capture the sales pattern illustrated by the solid curve in Figure 1 without separating different sales components.

The model we propose is termed as the Generalized Diffusion Model with Repeat Purchases (GDMR). We utilize a branch of mathematics named fractional calculus to develop this model. Specifically, the GDMR generalizes the fundamental differential equation governing the Bass Model, and employs a non-integer integral operator with flexible order, thereby rendering the Bass Model a special case of the extended model. The GDMR adopts an approach that is different from those in the prior literature in that it captures the sales growth rate using a non-integer order integral equation, rather than an integer-order differential equation as used in the prior literature.

Compared with existing models, the GDMR makes several important theoretical and methodological contributions to the existing literature on product diffusion and repeat purchases. First, developed by adding only one parameter to the classic Bass Model, the GDMR retains the Bass Model's parsimony as well as its insightful behavioral explanation concerning adopters' decisions in a diffusion process, which is a significant advantage compared to other benchmark models that do not generalize the Bass Model. Second, compatible with sales of technology products including both hardware and software products, the GDMR can count both replacement and multi-unit ownership purchases, further broadening its application compared to models that include only one of these repeat-purchase components. Third, the GDMR accounts for a continuum of sales scenarios ranging from a low boundary scenario in which no repeat purchase takes place to a top boundary scenario in which all adopters, on average, make one repeat purchase in each time-period, making the model applicable to a wide range of technology products. Fourth, marketing mix variables can be incorporated into the GDMR, thus further enhancing its potential in helping firms make better marketing decisions. Fifth, our empirical analysis indicates that the GDMR delivers superior performance in both model fitting and sales prediction compared to benchmark diffusion and repeat purchase models. Due to the challenges of developing a robust repeat purchase model, there has been limited advancement in this area in the last few decades. The demonstrated theoretical advantages and superior empirical performance of the GDMR makes a strong case that the proposed model represents a major advancement to the literature on product diffusion and repeat purchases.

DEFINING REPEAT PURCHASES

Before delving into model development, we first discuss the drivers of repeat purchases and a conceptual framework that helps explain the types of innovations and clarify what qualify as repeat purchases. In other words, we explain our unit of analysis that helps define the scope of adoptions and repeat purchases.

Repeat purchases may result from either product replacements (e.g., buying a new laptop computer to replace an old one) or multi-unit ownerships (e.g., one household owning multiple laptop or tablet computers). Drivers of replacement purchases vary from nondurable to durable products. For nondurable products, replacement usually occurs as the result of consumption (Kamakura & Balasubramanian, 1987). For durable products, replacement typically takes place when a product under consumption fails to meet the requirements of the user (Kamakura & Balasubramanian, 1987; Steffens, 2003). This failure may be caused by perceived or actual wear and tear as a result of consumption or by changes in a user's own needs or expectations that can only be met by a different version or generation of the original product. In high-tech markets, in particular, replacement is most often driven by product upgrade releases instead of wear and tear, as these markets experience frequent changes in the forms of quality improvements and/or price decreases (Gordon, 2009).

Motivations behind multi-unit adoptions may vary (Steffens, 2003). Common reasons include, for example, using the product in different locations (e.g., TV sets for different rooms of a house, desktop computers for office and home uses), providing extra capacity for peak demand (e.g., multiple photocopiers at the same location, multiple external data storage drives), or meeting the demand for different functionalities that require multiple versions of a product (e.g., a light laptop computer for traveling and a gaming laptop for entertainment, different versions of a software on different machines to deal with compatibility issues).

We next draw on the typology of product innovation by Henderson and Clark (1990) to define our unit of analysis — a family of products for which adoptions and repeat purchases take place. Henderson and Clark (1990) classify product innovations based on the amount of corresponding changes to the product's core components and product architecture. For example, a product innovation is considered radical if it introduces a new set of designs to the core components as well as a new architecture that links the core components. If the design of the core components and the architecture undergo only small improvements, the innovation is deemed incremental. In this study, we count a new purchase as a repeat purchase only if the newly purchased product is the same as the original product or a variation of the original product resulting from an incremental innovation. In other words, our unit of analysis is a line of products/innovations that differ incrementally. Within such a unit of analysis, the first purchase made for any of the products in the product line is considered the initial purchase, substituting a product by another in the same product line constitutes replacement, and simultaneous ownership of multiple units of the product line represents multi-unit ownership.

As technology advances and different products emerge based on innovation types other than incremental innovation, adopters may leave a focal product line and buy new products, which can lead to the obsolescence of the focal product line over time. When they buy such new products, even if they have similar functionalities, the purchases can no longer be counted as repeat purchases; instead, they are departures from the focal product line. For example, we can count the purchases of new models of DVD players as repeat purchases, but we need to treat upgrades to Blu-ray players as departures. While replacements and multi-unit ownerships increase repeat purchases, departures decrease repeat purchases. The goal of this study is to develop a unified model that can capture adoptions and different forms of repeat purchases, as well as the declining rate of repeat purchase due to departures and obsolescence.

GENERALIZED DIFFUSION MODEL WITH REPEAT PURCHASES

Developing mathematical models to capture a diffusion of innovation process has been a central part of the diffusion of innovation research. Pioneering scholars have presented various models to estimate and predict diffusion of innovation in a population. A comprehensive review of these models has been provided by Meade and Islam (1998) and Mahajan et al. (2000). Applying a bell-shaped growth curve to capture a noncumulative diffusion process is at the center of these endeavors. A mathematical realization of the bell-shaped diffusion process is represented by the Bass Model (Bass, 1969), which forms the foundation of the model we propose in this study.

The Bass Model stipulates that the noncumulative rate of adoptions of a product at time t, denoted by y(t), and the cumulative number of adoptions, denoted by Y(t), satisfy the following first-order differential equation:

$$\frac{dY(t)}{dt} = y(t) = pm + (q-p)Y(t) - \frac{q}{m}Y^2(t), t \geq 0 \tag{1}$$

where *p* and *q* are coefficient of innovation and coefficient of imitation, respectively, and m represents the size of the market potential. Note that Eq. (1) is an integer-order differential equation. Solving this equation with the initial condition $Y(0) = 0$ yields:

$$Y(t) = \frac{m(1-e^{-(p+q)t})}{(1+\frac{q}{p}e^{-(p+q)t})}, \tag{2}$$

$$y(t) = \frac{m(p+q)^2}{p}\frac{e^{-(p+q)t}}{(1+\frac{q}{p}e^{-(p+q)t})^2}. \tag{3}$$

The noncumulative version of the Bass diffusion curve, y(t), effectively captures the bell-shaped growth pattern of adoptions, whereas the cumulative version Y(t) exhibits the well-known S-shaped curve.

Many of the newer diffusion models have generalized earlier works. Model extensions in this literature have mostly concentrated on revising the right-hand side of the growth differential equation represented by Eq. (1). Adopting a different modeling approach, we present a generalization by extending both the left-hand side and the right-hand side of the diffusion rate differential equation. In particular, we draw upon a branch of mathematics called fractional calculus, which generalizes differentiation and integration so that non-integer-order differential and integral operators become possible. Interested readers can refer to Samko et al. (1993), Podlubny (1999), Kilbas et al. (2006), and Baleanu et al. (2012) for comprehensive reviews of the fractional calculus literature.

Built on fractional calculus and the Bass Model, our model incorporates repeat purchases into the Bass Model by formulating product sales as a fractional integral of its underlying diffusion process. To yield sales, the new model integrates the adoptions at the present time with the fraction of existing adopters who make repeat purchases at the present time. We call the extension model the Generalized Diffusion Model with Repeat Purchases or GDMR. The GDMR generates an asymmetrical bell-shaped sales growth curve as illustrated in Figure 1. The detailed development of the model is presented in the next subsection.

Integrating Adoptions and Repeat Purchases

The GDMR assumes that a noncumulative sales trend, S(*t*), is governed by the following fractional integral equation:

$$S(t) = I^{\beta}y(t) = I^{\beta}(\, pm + (q-p)Y(t) - \frac{q}{m}Y^2(t)),\ t \geq 0, \tag{4}$$

in which Y, y, p, q, and m represent cumulative adoptions, noncumulative adoptions, coefficient of innovation, coefficient of imitation, and market potential, respectively. β ($0 \leqslant \beta \leqslant 1$), which we term the coefficient of repeat purchases, is a new parameter added to the Bass Model, and its value determines the proportion of repeat purchases. I^{β} is a non-integer fractional integral of order β. We next explain the rational for Eq. (4) and the role of the newly added coefficient of repeat purchases.

In the presence of repeat purchases, sales are composed of adoptions (first-time purchases) and repeat purchases. We assume that a consumer's adoption and repeat purchases do not occur at the same time. For instance, a repeat purchase made at time *t* corresponds to an adoption that has taken place prior to time t, i.e., in [0, *t*). Therefore, the rate of repeat purchases at t depends on the cumulative number of existing adoptions in [0, *t*).

The rate of repeat purchases at time *t* depends on the rate at which adopters from $\tau \in [0, t)$ make repeat purchases at *t*. To understand this factor, we first consider a high repeat purchase scenario in which the average frequency of repeat purchase equals 1, implying that, on average, each existing adopter makes one repeat purchase in each unit time. We call this scenario the periodic-repeat-purchases. A good example of this scenario is that consumers renew their subscription to a software product every year. Under this scenario, the rate of repeat purchases at time *t* equals the cumulative number of adoptions just before time t. This number is added to the instantaneous rate of adoption exactly at time *t* to produce the sales rate at *t*. This means that the sales rate at time t equals the cumulative number of adoptions up to and including time *t*. Mathematically, this entails that, under the periodic-repeat-purchases scenario, sales at time *t* is a first-order integration of adoptions in [0, *t*]. Formally, we have

$$S(t) = \int_0^t y(\tau)\, d\tau = Y(t) = I^{\beta=1} y(t). \tag{5}$$

Therefore, under this periodic-repeat-purchases scenario, the noncumulative sales rate at time *t* equals the cumulative number of adoptions at time t, hence following an S-shaped curve. Note that the coefficient of repeat purchases (β) in Eq. (5) and the average frequency of repeat purchases both equal 1 in this scenario.

We next examine another boundary scenario, in which no repeat purchase occurs, hence sales are composed of only adoptions. We call this *adoption-only* scenario, under which the sales rate follows the noncumulative adoption rate defined by the Bass Model, which is essentially Eq. (4) with β = 0:

$$S(t) = y(t) = I^{\beta=0} y(t). \tag{6}$$

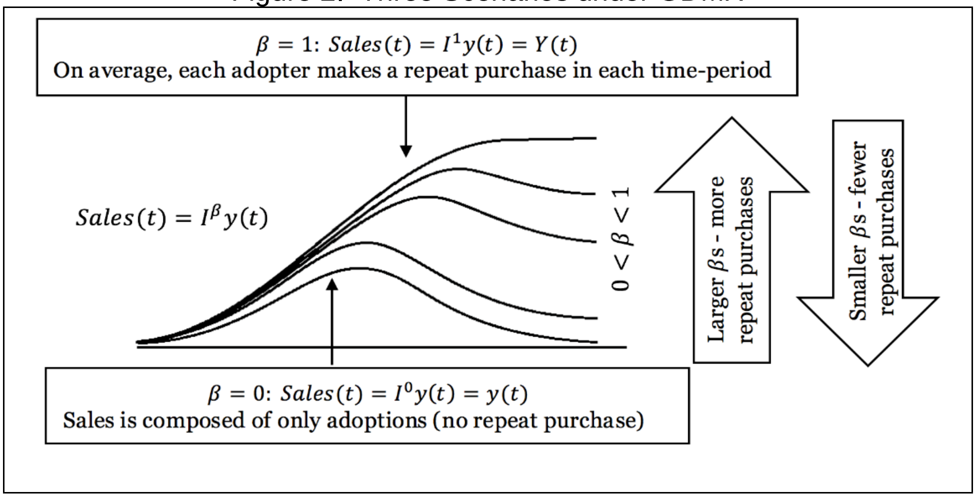
In this case, the sales rate follows a symmetrical bell-shaped curve. Here, the coefficient of repeat purchases (β) and the frequency of repeat purchases both equal 0.

In most practical scenarios, the average frequency of repeat purchases likely falls between the two boundary scenarios described above [In rare cases where the average frequency of repeat purchases is greater than 1, we can redefine the unit of time so that the frequency falls below 1]. Specifically, existing adopters, on average, make *more than zero but less than one* repeat purchases per unit time. Integrating such a rate of repeat purchases into the instantaneous adoption rate yields a sales curve that falls between the S-shaped curve (with β =1) and the symmetrical bell-shaped curve (with β =0). Mathematically, we use the *fractional integral* of noncumulative adoption shown in Eq. (4), with $0 < \beta < 1$, to represent sales rate under the *intermediate* scenario. In this scenario, the coefficient of repeat purchases (β) increases with the frequency of repeat purchases.

For the intermediate scenario in which the order of integral operator in Eq. (4) is between 0 and 1, the exact rate at which adoptions from $\tau \in [0, t)$ result in repeat purchases at t depends on the formulation for the fractional integral we use to operationalize the GDMR. We elaborate on our choice of fractional integral for the GDMR in the next subsection.

The dynamics of the GDMR under the three discussed scenarios are illustrated in Figure 2. In this figure, the bottom bell-shaped curve represents the adoption-only scenario ($\beta = 0$); the top S-shaped curve depicts the periodic-repeat-purchases scenario ($\beta = 1$). In between the two boundary curves are intermediate curves corresponding to $0 < \beta < 1$. Increasing the value of β causes the sales curve to shift away from the symmetrical bell-shaped curve corresponding to $\beta = 0$ and move closer to the S-shaped curve corresponding to $\beta = 1$, implying a higher sales rate due to more repeat purchases.

Figure 2. Three Scenarios under GDMR



Built on fractional integral, Eq. (4) defines the sale rate with repeat purchases. The rate of repeat purchases at time t, $R(t)$, is the difference between sales rate and adoption rate at t:

$$R(t) = S(t) - y(t) = I^\beta y(t) - y(t). \quad (7)$$

As shown in Eqs. (1) and (3), the adoption rate, as defined in the Bass Model, depends on the adoption parameters p, q, and m, and does not change with β. The sales rate $S(t)$, as shown in Figure 2, increases with β. Therefore, the rate of repeat purchases defined in Eq. (7) is an increasing function of β.

Model Operationalization

In the previous section, we have explained that under the periodic-repeat-purchases scenario, on average all adopters make a repeat purchase in each time-period, whereas in the adoption only scenario, adopters make no repeat purchase. Under the intermediate scenario, to determine the exact rate at which adoptions from $\tau \in [0, t)$ result in repeat purchases at t, we need to determine the formulation for the fractional integral in Eq. (4). We use the Reimann-

Liouville integral (Kilbas et al. 2006) to operationalize the GDMR [The GDMR as presented in Eq. (4) is a general model and can be operationalized using different fractional integral operators. We select the Riemann-Liouville fractional integral based on its performance in our empirical analysis]. The Riemann-Liouville fractional integral of order $\theta > 0$ of a function *f* is defined as:

$$I^{\theta} f(t) := \int_0^t \frac{1}{\Gamma(\theta)} (t-\tau)^{\theta-1} f(\tau) d\tau, \quad \theta > 0, \; t > 0, \tag{8}$$

and $\Gamma(x)$, the Gamma function, for $x \in \mathbb{R}$, $x > 0$, is

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt. \tag{9}$$

Thus, we can also present the GDMR as:

$$S(t) = I^{\beta} y(t) = \int_0^t \frac{1}{\Gamma(\beta)} (t-\tau)^{(\beta-1)} y(\tau) d\tau, \quad t > 0. \tag{10}$$

Next, we elaborate on the rate of repeat purchases by adopters based on Eq. (10).

In an important recent study, Tarasov (2018) proposes that fractional calculus can capture an *economic process with memory*, in which memory is defined as the dependence of an endogenous (predicted) variable at the present time on the history of the changes of an exogenous (predictor) variable in a time frame. Following Tarasov's memory concept, we can interpret that Eq. (10) represents the sales of a product as an economic process with a memory of adoptions from the past that generate repeat purchases at the present time.

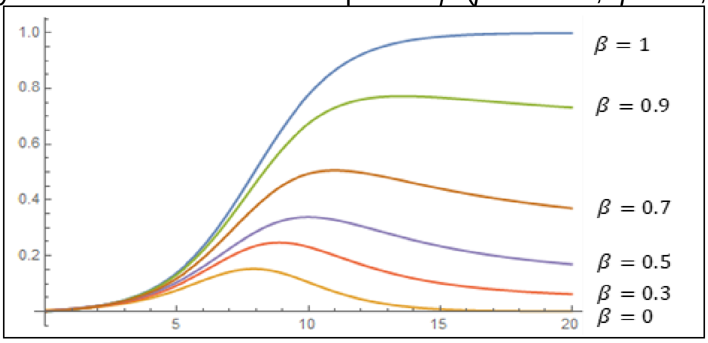
Approximation and Dynamics of GDMR

Estimating parameters based on Eq. (10) may present difficulties due to the complexity of computation associated with the fractional integral operator. Therefore, we need a mathematical operator that has desirable properties and is computationally feasible. We introduce the operator $I_{n,k}^{\beta}$ instead of I^{β} and reformulate the GDMR as:

$$S(t) = I_{n,k}^{\beta} y(t), \tag{11}$$

where *n* and *k* are parameters of the approximate operator. Increasing the values of *n* and *k* results in $I_{n,k}^{\beta}$ converging to I^{β} [The detailed derivation of the approximate operator and its convergence to the original one are available from the authors upon request].

Figure 3. Dynamics of GDMR with Respect to β (*p* = 0.005, *q* = 0.6, and *m* = 1)



Based on the operationalization presented above, we conduct numerical analysis to examine how the amount of repeat purchases changes with the coefficient of repeat purchases (β). The results are summarized in Figure 3, which illustrates how the GDMR sales curve varies with β (the values of the other parameters are fixed at *p* = 0.005, *q* = 0.6, and *m* = 1). This figure provides a clearer picture regarding how the sales curve changes with the value of the coefficient of repeat purchases (β). Consistent with our previous theoretical predictions, a higher β value in the GDMR is linked to a higher sales curve, thus representing a higher level of repeat purchases. As the value of β increases from 0 to 1, the sales curve moves further away from the bell-shaped curve with no repeat purchase and closer to the S-shaped curve representing on average one repeat purchase per unit-time by existing adopters.

Incorporating Marketing Mix Variables

It is well understood that marketing mix variables (e.g., price and advertising) can affect a diffusion process. Likewise, it is expected that marketing mix variables can influence the magnitude of repeat purchases and subsequently that of total sales. In this section, we examine how marketing mix variables can be incorporated into the GDMR.

Prior research has explored various ways of accounting for the influence of marketing mix variables on the diffusion of products (Bass et al. 2000). Here, we implement the approach used in the Generalized Bass Model (GBM) (Bass et al. 1994). Specifically, we incorporate the marketing mix variables into cumulative sales by the GDMR as:

$$CumS(t) = I^{\beta+1}[y(X(t))], \tag{12}$$

where *CumS*(*t*) is cumulative sales, *y* is periodic adoptions, and $X(t)$ is the *cumulative marketing effort* as defined in the GBM by Bass et al. (1994). For instance, including price (*Pr*) and advertising (*Adv*), $X(t)$ takes the following form:

$$X(t) = t + \gamma * ln\left[\frac{Pr(t)}{Pr(0)}\right] + \delta * ln\left[\frac{ADV(t)}{ADV(0)}\right], \tag{13}$$

where γ and δ are coefficients capturing the effects of changes in price and advertising, respectively. To differentiate it from the GDMR without marketing mix variables, we name the version with marketing mix variables the *Generalized Diffusion Model with Repeat purchases and Marketing Mix Variables* (GDMRX). GDMRX in its noncumulative form can be shown as:

$$S(t) = x(t) * I^{\beta}[y(X(t))], \tag{14}$$

where *x*(*t*) is *noncumulative marketing effort* as defined in the GBM.

PERFORMANCE EVALUATION

In this section, we empirically evaluate the model fitting and the prediction accuracy of the GDMR compared to alternative models [Sample source code for the empirical analysis is available from the authors upon request].

Benchmark Models

Because the GDMR captures all components of sales including adoptions, replacement purchases, and multi-unit purchases, and all three sales components are expected to be present in our datasets, we need to compare the GDMR with alternative models that also include adoptions, replacement purchases, and multi-unit purchases. However, unlike the broader literature on product adoption, the literature concerning repeat purchases is rather limited, and we could not find readily available models that explicitly incorporate the three components of sales. This is likely due to the significant challenges in modeling all commonly seen repeat purchase scenarios. Therefore, we have to develop our own benchmarks based on models proposed in the prior literature.

For comparison purposes, we consider two benchmark sales models that incorporate initial purchases, replacements, and multi-unit ownerships. The first model incorporates the Bass Model for initial purchases, the Kamakura and Balasubramanian (KB) (1987) replacements model, and the Bayus et al. (BHL) (1989) multi-unit purchases model. For expositional convenience, we name this model the *Bass-KB-BHL Sales Model*. The second benchmark model is composed of the same initial purchases model and replacement model, but uses a different multi-unit purchase model developed by Steffens (2003). We name this model the *Bass-KB-Steffens Sales Model*. To help readers better understand the two benchmark models, we next briefly elaborate on the KB Model, BHL Model, and Steffens Model.

Kamakura and Balasubramanian (1987) formulate a product replacement model for durable products based on two assumptions: (i) a product is immediately replaced after it fails to perform up to users' expectations, and (ii) these failures can be represented by a probability distribution function over all product units. The model can be expressed as:

$$r(t) = \sum_{i=1}^{t-1}[y(i) + r(i)]\,[Sur(t-i-1) - Sur(t-i)], \tag{15}$$

where [$y(i)$ + $r(i)$] represents sales at time i and is composed of initial purchases, $y(i)$, and product replacements, $r(i)$ [The model assumes that multi-unit ownership purchases are insignificant]. $Sur(\tau)$ is a survival function capturing the probability that a product unit fails after τ. For example, with truncated normal distribution as the distribution of survivals [There is extensive support in the literature for using the truncated normal distribution for this purpose (Steffens 2003)], $Sur(\tau)$ takes the following form:

$$Sur(\tau) = \frac{\Phi(\frac{w\tau}{L} - h)}{\Phi(-h)}, \tag{16}$$

where $w = h + \phi(-h) / \Phi(-h)$, $\phi(.)$ is the standard normal probability density function, and

$$\Phi(x) = \int_{z=x}^{\infty} \phi(z)dz. \tag{17}$$

Bayus et al. (1989) develop a model that explicitly considers multi-unit ownerships based on the premise that the older is the product unit in use, the more likely is the purchase of an additional unit. The proposed functional form is

$$mul(t) = \sum_{i=1}^{t-1} siu_{i,t-1}\, g(t-i), \tag{18}$$

where *mul*(t) represents the number of multi-unit purchases at time t, $g(\tau)$ is the hazard rate for multi-unit purchases at τ, and

$$siu_{i,t} = [y(i) + r(i) + mul(i)]Sur(t-i). \tag{19}$$

In this model, $g(\tau)$ is empirically determined by Bayus et al. (1989). Following Steffens (2003), we adopt a logistic growth function for $g(\tau)$:

$$g(\tau) = \frac{\delta(1-e^{-\alpha\tau})}{(1+\beta e^{-\alpha\tau})}. \tag{20}$$

Steffens (2003) views the purchase of an additional unit as a diffusion process and formulates it based on the Bass Model:

$$\frac{\mathrm{d}mult(t)}{\mathrm{d}t} = (\pi Y(t) - mult(t))(a + b * mult(t)), \tag{21}$$

where *mult*(t) is the cumulative number of first additional units purchased by users, $Y(t)$ is the cumulative adoptions representing the upper limit for multi-unit ownerships, a represents the external influences on first multi-unit adoptions, and b represents the word-of-mouth influences on first multi-unit adoptions. Steffens suggests that the adoption of more additional units by users (e.g., second additional units, third additional units) can be similarly modeled. The Steffens Model does not have a closed-form solution and can only be numerically integrated.

By summing up the aforementioned models for initial, replacement, and multi-unit ownership purchases, we obtain two benchmark models: Bass-KB-BHL and Bass-KB-Steffens. We empirically compare the GDMR against Bass-KB-BHL and Bass-KB-Steffens.

Model Fitting and N-Period-Ahead Prediction

We use three aggregate sales datasets to evaluate the model fitting and N-period-ahead prediction accuracy of the GDMR and the two benchmarks. The datasets include annual sales of Notebook Computers from year 2005 to 2014 (Morgan Stanley 2015), PC total global annual sales from year 2006 to 2015 (Statista 2014), and iPad sales from the third quarter of year 2010 to the second quarter of 2017 [Data obtained from Apple's quarterly summaries. For instance, 2013 Q4 data is found in https://www.apple.com/newsroom/pdfs/ q4fy13datasum.pdf. Starting from the third quarter of 2010, we sum four quarters of sales to derive one year worth of iPad sales]. All three datasets include only aggregate sales data, and a breakdown into adoption purchases, replacement purchases, and additional-unit purchases is unavailable.

Following the parameter estimation approach by Srinivasan and Mason (1986), we use the changes in the cumulative number of adoptions between two consecutive periods to represent the noncumulative adoptions trend in the benchmark models.

Table 1. GDMR Parameter Estimation for Notebook Computer Sales

	Estimate	Standard Error	t-statistics	P-value
β	0.46	0.110	4.9	0.00
p	0.019	0.005	4.5	0.01
q	0.627	0.108	5.7	0.00
m	4.6×10^8	1.5×10^8	3.2	0.02
s_0	4.8×10^7	9.9×10^6	4.9	0.00

For the first two datasets, due to potential left-hand truncation, we add an intercept, s_0, to both the GDMR and the benchmark models to help improve model estimation. In our application of the KB replacement model, we assume that sales include additional-unit purchases as well (i.e. sales in Eq. (15) is of the form [*y*(*i*) + *r*(*i*) + *mul*(*i*)] instead of [*y*(*i*) + *r*(*i*)]). Because our test data is for consumer electronics, to guide the estimation of the benchmarks' replacement component, we assume that the average life of the products is less than or equal to eight years.

The GDMR parameter estimates for the three products are summarized in Tables 1, 2, and 3, respectively. As we can see from the tables, except for parameter *p* for PC sales, all other parameter estimates for all three products are statistically significant, indicating a good overall model fit in all three cases. The value of the coefficient of repeat purchases (β) ranges from 0.24 to 0.48 indicating that sales curves for these three products lie between the bell-shaped and S-shaped curves. Therefore, the traditional diffusion parameter values for *p*, *q*, and *m* would have been biased without explicitly incorporating repeat purchases in the model formulation. Furthermore, the estimate for s_0 is statistically significant for the first two products, showing that it is an effective way to addressing the left-hand data truncation issue.

We also compare the GDMR against the two benchmark models using the three datasets; their model fitting and prediction performance is summarized in Table 4. The fitting accuracy is measured in terms of R^2 and MAPE (mean absolute percentage error). From the summary, it is clear that the GDMR, despite having fewer parameters, consistently leads to better fits than the two benchmark models do.

Table 2. GDMR Parameter Estimation for PC Sales

	Estimate	Standard Error	t-Statistics	P-value
β	0.24	0.130	5.70	0.00
p	0.014	0.009	1.56	0.18
q	0.725	0.201	3.60	0.02
m	4.8×10^8	2.0×10^8	2.35	0.07
s_0	2.3×10^8	1.7×10^7	13.44	0.00

Table 3. GDMR Parameter Estimation for iPad Sales

	Estimate	Standard Error	t-Statistics	P-value
β	0.48	0.018	28.44	0.00
p	0.05	0.003	18.18	0.00
q	1.17	0.046	25.36	0.00
m	1.53×10^8	5.5×10^6	27.70	0.00

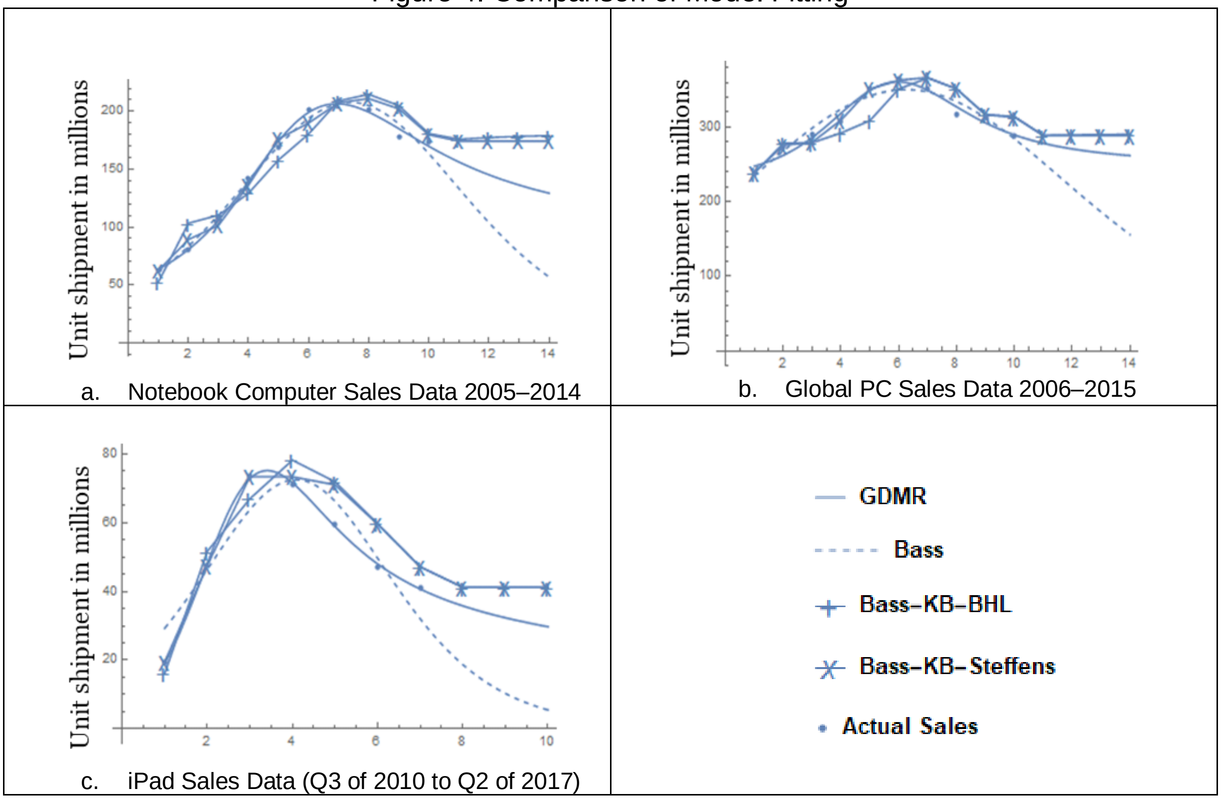
The prediction performance of the three models is evaluated based on the MAPE of one-year-ahead and two-years-ahead sales prediction. The results in Table 4 show that, with the exception of one-year-ahead prediction for notebook computers and two-years-ahead prediction for PCs (for which Bass-KB-Steffens and Bass-KB-BHL respectively are the best models), the prediction accuracy of the GDMR is clearly better than that of the benchmark models.

The model fitting results for the GDMR, Bass-KB-BHL, and Bass-KB-Steffens are depicted in Figures 4 for the three datasets. To contrast these repeat purchase models with the adoption-only Bass Model, we include the fitted Bass diffusion curve in these figures as well. For iPad sales, the GDMR appears to fit the data better than the benchmarks after the sales process peaks. This can be attributed to the fact that the benchmark models overestimate repeat purchases rate in the later stages of product lifecycle. As expected, the Bass Model fits the data reasonably well before the peak point of sales, but its predicted sales curve declines much more quickly than the curves produced by the other three models do after the peak, resulting in poor model fits. This is particularly obvious in Figure 4 for iPad sales, where the fitted Bass Model curve is clearly off. Therefore, the Bass Model is not recommended when repeat purchases cause the aggregate sales to exhibit a clear asymmetric curve.

Table 4. Comparison of Model Fit and Prediction Accuracy for GDMR, Bass-KB-BHL, and Bass-KB-Steffens

		Full data fit		Prediction Accuracy	
				Years Ahead MAPE	
		R^2	MAPE	One	Two
	GDMR	0.9995	1.68	14.46	8.79
Notebook Computer	Bass-KB-BHL	0.9054	10.09	6.83	31.09
	Bass-KB-Steffens	0.9573	5.40	4.64	25.75
	GDMR	0.9995	2.08	0.94	8.84
PC	Bass-KB-BHL	0.6731	5.25	10.28	8.35
	Bass-KB-Steffens	0.8390	3.11	10.28	18.11
	GDMR	0.9999	1.34	3.10	1.90
iPad	Bass-KB-BHL	0.7763	15.31	14.55	37.84
	Bass-KB-Steffens	0.8439	9.05	14.48	37.62

Figure 4. Comparison of Model Fitting



a. Notebook Computer Sales Data 2005–2014

b. Global PC Sales Data 2006–2015

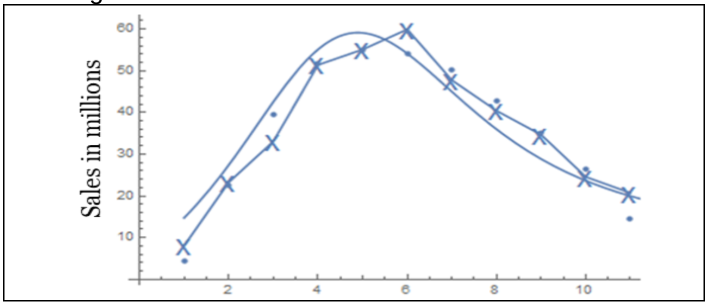
c. iPad Sales Data (Q3 of 2010 to Q2 of 2017)

With all results considered, we conclude that the GDMR is a better model for both model fitting and prediction of future sales.

Performance Evaluation with Marketing Mix Variables

As we show in Eq. (12), the GDMR can be extended to include the influence of marketing mix variables such as price and advertisement. Hence, we collect additional data to evaluate the performance of this extension model labeled as GDMRX. Since the two benchmark models do not capture the impact of marketing mix variables, we compare GDMRX with the baseline GDMR only.

Figure 5. GDMR and GDMRX for iPod Sales Data



We use iPod sales from 2004 to 2014, obtained from Apple's quarterly summaries, to illustrate the performance of the GDMRX. To obtain the average selling price of the iPod, we divide the revenue generated by iPod sales by the corresponding number of units sold. The comparison of model fitting between the GDMR and the GDMRX is shown in Figure 5. The parameter estimates and model fitting measures are summarized in Table 5.

These results show that both the GDMR and the GDMRX perform well on the iPod sales data with the GDMRX having a slight edge in terms of sum of squared errors (SSE). This is not surprising given that the GDMRX is a more flexible model than the GDMR.

Table 5. GDMR and GDMRX Parameter Estimates, R^2, and SSE for iPod Sales

	β	p	q	m	γ	R^2	SSE
GDMR	0.278 (0.066)	0.037 (0.004)	0.656 (0.066)	2.18×10^8 (0.33×10^8)	-	0.9998	9.8×10^{13}
GDMRX	0.21 (0.096)	0.027 (0.01)	0.603 (0.065)	2.56×10^8 (0.58×10^8)	–0.745 (0.787)	0.9998	8.4×10^{13}

Note: SSE, sum of squared errors; values in parentheses represent standard errors. Estimation results are based on fitting the cumulative forms of the GDMR and GDMRX to cumulative sales data.

CONCLUSIONS AND DISCUSSIONS

In this section, we provide a brief summary of the proposed model, highlight its methodological and theoretical contributions, and discuss its practical implications.

Summary

This study proposes a robust sales growth model to predict lifecycle sales of technology products in the presence of frequent repeat purchases. Based on a branch of mathematics called fractional calculus, we develop a novel sales model, the *Generalized Diffusion Model with Repeat Purchases* (GDMR), to account for repeat purchases that have become increasingly prevalent in today's market of technology products including both hardware and software products. The GDMR generalizes the classic Bass Model using an integral operator with fractional order, and produces sales growth curves that stay above the adoption trend. By assigning different values to the newly introduced parameter named *coefficient of repeat purchases*, the proposed GDMR can cover a wide continuum of sales growth scenarios ranging from symmetrical bell-shaped sales curves to S-shaped sales curves. This represents sales for a wide variety of products ranging from those that experience almost no wear and tear and offer no incentive for repeat purchases, to products that have a short lifetime, multiple versions, or a high rate of product upgrade releases. Our empirical tests show that the GDMR outperforms benchmark repeat purchase models.

Methodological and Theoretical Contributions

As explained earlier, because of the significant challenges in modeling repeat purchases, there have been limited development in this area in the last few decades. Based on our extensive literature review, no prior research has considered all commonly seen repeat purchase scenarios in a single model. By proposing a robust repeat purchase model to fill this void, this research represents an important addition to the literature on product diffusion and repeat purchases. The significant theoretical and methodological contributions made by this research are demonstrated through the major advantages of the GDMR compared to alternative models:

- As a direct extension to the Bass Model, the GDMR retains the key advantages of the Bass Model, including its parsimony and interpretability, two of the critical advantages that cement the Bass Model's status as the most widely adopted diffusion model, while alternative models are more cumbersome in nature.
- The GDMR is a theoretically superior model than existing repeat purchase models because it considers both replacement purchases and multi-unit purchases and covers a wider continuum of repeat purchase scenarios ranging from zero repeat purchases to highly frequent repeat purchases, making it applicable to a wide variety of products.
- Unlike alternative repeat purchase models, the GDMR can incorporate the influence of marketing mix variables, thus further enhancing its potential in helping firms make better marketing decisions.
- The GDMR outperforms benchmark repeat purchase models.

The contribution of the present research is not limited to the literature on product diffusion and repeat purchases. By accounting for repeat purchases using a fractional integral of the adoptions trend, the GDMR presents a new interpretation of fractional integral, thereby contributing to the broad literature of fractional calculus in applied mathematics and science. It is very well known that integer-order integrals and derivatives have simple and clear geometric and physical interpretations, thus guiding numerous applications of these tools in science (Podlubny 2002). Interpreting fractional integrals and derivatives, however, has proven challenging. For more than 300 years since the introduction of fractional calculus, no clear geometric and physical interpretation of fractional derivatives and integrals has been introduced

(Podlubny 2002). In the first international conference on fractional calculus held in New Haven (USA) in the year 1974, discovering the physical and geometric interpretations of fractional calculus was included in the list of open problems (Ross 1975). Only in more recent years, different interpretations of fractional calculus have emerged, including the economic interpretation proposed by Tarasova and Tarasov (2017). By proposing an interpretation of fractional integral in the context of product adoptions and repeat purchases, this study represents the newest contribution to the cross-disciplinary endeavor of interpreting fractional calculus. To the best of our knowledge, the present research is the first in either the IS literature or the more general management sciences literature to apply fractional calculus in the study of business problems.

Practical Implications

In practice, sales data that counts only first-time purchases is rarely available. The vast majority of sales data maintained by firms or third parties includes both first-time and repeat purchases. Therefore, making a robust repeat purchase model available to firms has important and wide-ranging practical implications.

Some of practical implications of the GDMR have already been discussed in the previous subsection along with the model's theoretical advantages. Here we would like to point out several important practical implications brought by the nature and versatility of the GDMR. First, by accurately differentiating adoptions from repeat purchases, the GDMR can overcome parameter estimation bias resulting from force-fitting the Bass Model to sales data that is typically "contaminated" by repeat purchases, thus helping firms better understand the progress of product diffusions and make more accurate predictions of future sales. Second, the product diffusion literature has been centered around the seminal Bass Model (Bass 1969), based on which numerous extensions and applications have been proposed and analyzed, forming a vast body of knowledge that is invaluable to both researchers and businesses. Since it is a direct extension of the Bass Model, applying the GDMR allows firms to take advantage of this huge body of knowledge to support their operational and decision-making needs. Third, we have shown that the GDMR model can be extended to incorporate the impact of marketing mix variables such as pricing and advertising on the rate of repeat purchases. This important feature can help first assess the return of past spending on price discounts and market promotions and make recommendations on how to improve their future marketing operations.

REFERENCES

Baleanu, D., Diethelm, K., Scalas, E., & Trujillo, J. J. (2012). *Fractional calculus: models and numerical methods* (Vol. 3). World Scientific.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management science*, 15(5), 215-227.

Bass, F. M. (2004). Comments on "a new product growth for model consumer durables the bass model". *Management science*, 50(12_supplement), 1833-1840.

Bass, F. M., Jain, D., & Krishnan, T. (2000). Modeling the marketing-mix influence in new-product diffusion. *New-product diffusion models*, 99-122.

Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass model fits without decision variables. *Marketing science*, 13(3), 203-223.

Bayus, B. L., Hong, S., & Labe Jr, R. P. (1989). Developing and using forecasting models of consumer durables: The case of color television. *Journal of Product Innovation Management*, 6(1), 5-19.

Bhagat, R., Muralidharan, S., Lobzhanidze, A., & Vishwanath, S. (2018, July). Buy it again: Modeling repeat purchase recommendations. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 62-70).

Dodson Jr, J. A., & Muller, E. (1978). Models of new product diffusion through advertising and word-of-mouth. *Management Science*, 24(15), 1568-1578.

Gordon, B. R. (2009). A dynamic model of consumer replacement cycles in the PC processor industry. *Marketing Science*, 28(5), 846-867.

Gurbaxani, V., & Mendelson, H. (1990). An integrative model of information systems spending growth. *Information Systems Research*, 1(1), 23-46.

Guo, Z., & Chen, J. (2018). Multigeneration product diffusion in the presence of strategic consumers. *Information Systems Research*, 29(1), 206-224.

Hann, I. H., Koh, B., & Niculescu, M. F. (2016). The double-edged sword of backward compatibility: The adoption of multigenerational platforms in the presence of intergenerational services. *Information Systems Research*, 27(1), 112-130.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

Hu, Q., Saunders, C., & Gebelt, M. (1997). Diffusion of information systems outsourcing: A reevaluation of influence sources. *Information Systems Research*, 8(3), 288-301.

Jiang, Z., Qu, X. S., & Jain, D. C. (2019). Optimal market entry timing for successive generations of technological innovations. *MIS Quarterly,* 43(3), 787-806.

Jiang, Z., & Sarkar, S. (2009). Speed matters: The role of free software offer in software diffusion. *Journal of Management Information Systems*, 26(3), 207-240.

Kamakura, W. A., & Ealasubramanian, S. K. (1987). Long-term forecasting with innovation diffusion models: The impact of replacement purchases. *Journal of forecasting*, 6(1), 1-19.

Kilbas, A. A., Srivastava, H. M., & Trujillo, J. J. (2006). *Theory and applications of fractional differential equations* (Vol. 204). Elsevier.

Lee, C. Y., & Huh, S. Y. (2017). Technology forecasting using a diffusion model incorporating replacement purchases. *Sustainability*, 9(6), 1038.

Lilien, G. L., Rao, A. G., & Kalish, S. (1981). Bayesian estimation and control of detailing effort in a repeat purchase diffusion environment. *Management science*, 27(5), 493-506.

Loh, L., & Venkatraman, N. (1992). Diffusion of information technology outsourcing: influence sources and the Kodak effect. *Information systems research*, 3(4), 334-358.
Mahajan, V., Muller, E., & Wind, Y. (Eds.). (2000). *New-product diffusion models* (Vol. 11). Springer Science & Business Media.

Meade, N., & Islam, T. (1998). Technological forecasting—Model selection, model stability, and combining models. *Management science*, 44(8), 1115-1130.

Morgan Stanley. (2015). *Lowering Smartphone Forecast on Weaker China in 1Q15*.

Niculescu, M. F., & Whang, S. (2012). Research Note—Codiffusion of Wireless Voice and Data Services: An Empirical Analysis of the Japanese Mobile Telecommunications Market. *Information Systems Research*, 23(1), 260-279.

Olson, J., & Choi, S. (1985). A product diffusion model incorporating repeat purchases. *Technological Forecasting and Social Change*, 27(4), 385-397.

Podlubny, I. (1999). *Fractional differential equations*, Academic Press.

Podlubny, I. (2002). Geometric and physical interpretation of fractional integration and fractional differentiation. *Fractional Calculus and Applied Analysis*, vol. 5, no. 4, 2002, pp. 367-386.

Rao, A. G., & Yamada, M. (1988). Forecasting with a repeat purchase diffusion model. *Management Science*, 34(6), 734-752.

Rogers, E. M. (2003). *The diffusion of innovation 5th edition*. New York: Simon and Schuster.

Ross B. (ed.) (1975). *Fractional Calculus and Its Applications,* Lecture Notes in Mathematics, vol. 457, Springer-Verlag, New York.

Samko, S.G., Kilbas, A.A. and Marichev, O.I. (1993). *Fractional Integrals and Derivatives-Theory and Applications*. Gordon and Breach. Linghorne, PA.

Shmueli, G. and Koppius, O.R., (2011). Predictive analytics in information systems research. *MIS quarterly*, 553-572.

Srinivasan, V., & Mason, C. H. (1986). Nonlinear least squares estimation of new product diffusion models. *Marketing science*, 5(2), 169-178.

Statista. (2014). Global PC Sales Fall to Eight-Year Low. Retrieved November 17, 2016, www.statista.com/chart/4231/global-pc-shipments/.

Steffens, P. R., & Balasubramanian, S. (1998). A study of the time varying mean replacement age of consumer durables. *Marketing science conference*, 10-13 July, 1998, Fontainbleau, France.

Steffens, P. R. (2003). A model of multiple-unit ownership as a diffusion process. *Technological Forecasting and social change*, 70(9), 901-917.

Tam, K. Y. (1996). Dynamic price elasticity and the diffusion of mainframe computing. Journal of Management Information Systems, 13(2), 163-183.

Tarasova, V. V., & Tarasov, V. E. (2017). Economic Interpretation of Fractional Derivatives. *arXiv preprint arXiv*:1712.09575.

Tarasov, V. (2018). Generalized memory: Fractional calculus approach. *Fractal and Fractional*, 2(4), 23.

DECISION SCIENCES INSTITUTE

Emotion and Information Disclosure on Social Media

Yasine Bensouda
Virginia Tech, Pamplin College of Business
880 West Campus Drive, Blacksburg, VA, 24061
Email: yasineb@vt.edu

Tabitha L. James
Virginia Tech, Pamplin College of Business
880 West Campus Drive, Blacksburg, VA, 24061
Email: tajames@vt.edu, (540)231-3163

Christine Abdalla Mikhaeil
IESEG School of Management, Univ. Lille, CNRS, UMR 9221
3, rue de la Digue, 59000 Lille, France
Email: c.abdallamikhaeil@ieseg.fr, +33(0)320545892

ABSTRACT

Social media platforms give individuals the opportunity to express their ideas, share their perspectives, and connect with friends and strangers while building a community. We study how people reveal information when they engage with others on online social media platforms such as Instagram, Facebook, Twitter, and Snapchat. We asked college students at a European business school to keep a diary of the social media posts that they made for at least three days. Our qualitative analysis of the diaries reveals that strong emotions enhanced by isolation can lead students to reveal information such as their location, political viewpoints, or interests.

KEYWORDS: Social Media, Privacy, Information disclosure, Emotion, Functional model of self-disclosure

INTRODUCTION

Social media platforms offer users the ability to express their perspectives and tell others about their day-to-day life. Today’s young adults share content, respond to other’s content, and interact with others through social media more than any demographic group. In fact, “at no other time in history have young people enjoyed such an opportunity to make themselves visible to, and heard by, diverse audiences” (Herring & Kapidzic, 2015, p. 146). Although young adults desire to express themselves across social media platforms, the information revealed jeopardizes their privacy. Through the images, videos, and text they post, social media users reveal information about their interests, locations, appearance, and much more. Such information can be accessed by both the well-intentioned and the malevolent with ease. Although social media platforms have been a constant for young adults growing in the digital age, they became the primary way to safely keep in touch with friends and family during the COVID-19 pandemic (Molla, 2021).

Emotion and information disclosure are often associated in psychology literature through the concept of catharsis that is “the release of tension or emotion that can be achieved through verbal disclosure of that emotion” (Omarzu, 2000, p. 175). People use social media to regulate emotion by sharing and discussing their positive and negative emotions with others (Vermeulen et al., 2018). Emotion can thus be a driver of people’s information disclosure decisions. We study users’ emotional context when they decide to post on social media to see how emotion is linked with motives for information disclosure and the information that is ultimately revealed.

Emotion can be defined as a mental state of readiness for action that promotes behavioral activation and helps prioritize and organize behaviors in ways that optimize individual adjustments to the demands of the environment (Bagozzi et al., 1999; Lazarus, 1991). Self-disclosure research indicates that people benefit from revealing experiences that are emotionally important (Harber & Cohen, 2005; Hemenover, 2003). The disclosure decision model suggests that situational cues make possible social rewards from information disclosure salient (Omarzu, 2000). We build on this knowledge to suggest that felt emotions may prompt information disclosure on social media by increasing the salience of a desired reward. For example, a person experiencing enthusiasm for a new movie may be socially rewarded by likes of a post on social media about the movie.

We conducted a diary study in which participants were asked to document the social media posts they made over the course of three days including the emotions they were feeling, the motives for their posts, the information disclosed, and the rewards obtained. A qualitative analysis of the diaries revealed three emotions that prompted the release of three types of information on social media. First, we found that feelings of nostalgia for travel and socialization before the pandemic resulted in people making social media posts that revealed location and personal history information. Second, enthusiasm drove people to share personal interests. Third, we found that anger drove people to make posts about politics and particular causes. Overall, our results show that emotion can lead people to reveal personal information on social media.

LITERATURE REVIEW

Our study draws from research on emotions and social media use. Moreover, we leverage the functional model of self-disclosure and the disclosure decision model as theoretical foundations for our study. In the sections that follow, we briefly introduce these foundations.

Emotions and Social Media

There are two ways people can use social media for emotional regulation (Vermeulen et al., 2018). First, users may browse social media attempting to change their current emotional state (Myrick, 2015; van Ingen et al., 2016). Second, users may post to share and talk about their emotions (Bazarova et al., 2015; Rimé, 2009; Vermeulen et al., 2018). The first use can be passive in nature; that is, users are consuming content without contributing. The second use is active in nature; that is, users are contributing content to the platform and actively engaging with others. Our focus is to study how emotions drive active social media use and the information that emotional use reveals.

Studies that have focused on sharing emotion on social media have found that, for example, different platforms are used to share different types of emotions (Vermeulen et al., 2018). In their study, adolescents reported sharing positive emotions more on Instagram, Facebook, and Snapchat, whereas Twitter and Messenger were used more often to share negative emotions.

Another study found that people who shared positive and personally relevant emotions on social media reported more satisfaction (Bazarova et al., 2015). They also found that it was important that others were responsive to the emotional posts, and that both the quality and quantity of the replies mattered. We build on this research by studying the emotions that prompt social media sharing, the rewards that motivated it, and the information revealed in the posts.

Functional Model of Self-Disclosure

Derlega and Grzelak (1979) proposed five motives for self-disclosure: self-expression, self-clarification, social validation, relationship development, and social control. Self-expression or relief of distress allows people to “release pent-up feelings” (Prager et al., 1989, p. 564) and involves talking about problems and emotions (Omarzu, 2000). Self-clarification or identity clarification involves “talking about one’s beliefs and opinions to clarify one’s own position” (Prager et al., 1989, p. 564). Social validation involves “eliciting feedback from others, or validating one’s self-concept” (Prager et al., 1989, p. 564) and is a means to increase social acceptance (Omarzu, 2000). Relationship development allows people to find out about others and develop relationships with them (Prager et al., 1989). Finally, social control involves trying to control or exploit others (Prager et al., 1989).

People desire to be liked and accepted by others, so in the absence of another motive, people often self-disclose to gain social acceptance (Omarzu, 2000). They will disclose information that they think makes them look favorable to others and that reflects the image they wish to present to others. Situational factors may prompt a different goal to become salient. We propose that emotion can serve as such a situational prompt.

The Disclosure Decision Model

The disclosure decision model describes the process of information disclosure by breaking it into three stages (Omarzu, 2000). In the first stage, the saliency of an information disclosure motive is determined, and if no motive is salient, no disclosure is made. Therefore, when a user does not post, there is no motive to do so which could indicate general apathy. To consider the motives that may be salient, Omarzu (2000) draws the five motives from the functional model of self-disclosure and suggests that situational cues and individual differences may influence the goal that becomes salient in a particular situation.

Once a disclosure motive is salient, the second stage of the disclosure decision process involves determining if disclosing is appropriate and to whom the information will be disclosed. If disclosure is deemed to be an appropriate strategy and a target for disclosure is determined, an information disclosure is made. For example, if a user feels enthusiastic about a movie coming out and wants to share that excitement with her friends for social validation, she may consider posting her thoughts on the upcoming movie along with the trailer on social media. In the third stage, the person decides what to disclose. Disclosure decisions in this stage reflect the cost-benefit tradeoff; that is, if the perceived risks of disclosing the information outweigh the benefits, the person may limit the information disclosed.

METHOD

Study participants were asked to keep a social media diary for at least three days. The format of the diary was specified. For each post, the social media platform to which the participant was making the post was noted. The participants were also asked to indicate, for each post, their emotional state at the time of the post, their motivation for the post, the information the post contained, and any responses they got to their post. Our data collection took place during the COVID-19 pandemic and thus this major event shaped the emotions driving disclosures.

The study participants were college students at a European business school in their fourth year of a five-year university program. The 43 participants ranged in age from 21 to 28. There were 25 males, 7 females, and 11 who preferred not to specify their gender. The participants made 125 total social media posts during the data collection period. The posts were made primarily on Instagram (55 posts), Facebook (24 posts), and Twitter (17 posts). However, posts were also made on LinkedIn (10), Reddit (6), Snapchat (9), TikTok (1), and YouTube (3). The participants in our study thus used a variety of social media platforms during the data collection period. Of the 43 participants, 38 were active users who made at least one post during the data collection period. Five participants were passive users, i.e., they did not post to social media.

The diaries were analyzed using NVivo 12. We followed the coding procedures described in grounded theory (Strauss & Corbin, 1990). First, we performed open coding to identify the poster's emotional state, the motivations for the post, and the types of information disclosed. One researcher sought initial patterns in the diaries and debriefed the initial codes with two other researchers. We first noticed the absence of emotions from passive users and set them apart from active users. We found evidence of nostalgia, enthusiasm, and anger driving disclosure behaviors on social media. We also observed that the COVID-19 context heightened these emotions.

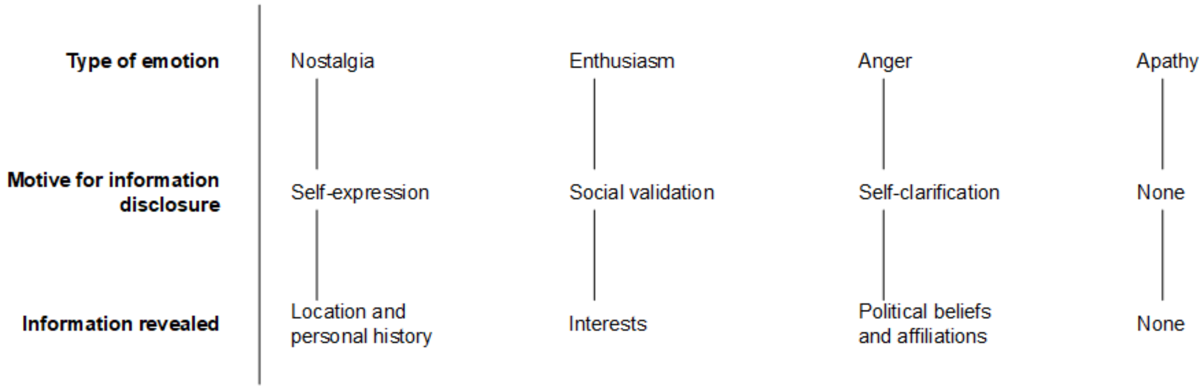
After our initial open coding, we followed an abductive approach iterating between data and theory using axial coding. For instance, we reviewed the literature on emotions and social media, which led us to reconsider our treatment of passive user diaries and integrate the role of apathy (i.e., defined as the absence of emotions) in our study. We used the disclosure decision model integrating apathy as a framework for our analysis (Omarzu, 2000).

Finally, we sought to refine the relationships observed in the disclosure behaviors. We crossed-checked our understanding of information disclosure behaviors on social media by using the functional model of self-disclosure to examine the relationships between type of emotion, motives for information disclosure and information revealed (Figure 1).

RESULTS AND DISCUSSION

Our main findings are summarized in Figure 1. Some of our participants considered themselves to be casual users of social media. These participants expressed apathy toward social media and reported at most passive use. Such people do not have an information disclosure goal that is salient for social media and therefore do not reveal information on it. These participants are represented on the far right of Figure 1. Our results indicate that emotions drive users to post and that different emotions lead them to reveal different types of information for different primary reasons. We found three different emotions to be frequently salient in our posters. Each of these emotions were associated with a different primary motive for making a post to social media and ultimately led to different types of information to be revealed about the poster. The emotion, the motive for information disclosure associated with that emotion, and the information revealed in the subsequent post are shown in Figure 1. We also provide sample quotes for each emotion from our data in Table 1. In the following sections, we discuss each emotion and the resulting information disclosures.

Figure 1. Summary of Main Findings



Nostalgia and Location

Nostalgia can be defined as a sentimental longing for the past and is "often identified as an emotional response brought on by dissatisfaction or detachment in the present, and anxiety for the future" (Gammon & Ramshaw, 2020, p. 132). The COVID-19 pandemic has led to extreme isolation and wistful thinking about happier times with friends in the past (Seattle Times Staff, 2020). Isolation and separation have given individuals the opportunity to reflect on happier moments before the outbreak. The ability to "look back and find comfort in pre-pandemic times has become a welcomed distraction" from isolation (Gammon & Ramshaw, 2020, p. 131).

We found evidence of nostalgia driving social media activity in our data. Our participants posted on social media to cope with the burden of the pandemic by re-living happier past events through their posts. The primary motive for such posts is to relieve distress (i.e., self-expression). Participants posted their previous experiences and vacations to show other users "happier times" before the pandemic. One participant revealed the motive for their post: "I think that the time period has something to do with that (Pandemic, Lockdown…), and to post something on social media reminds me of the time when we were free to go on vacation or party together" (L14). Our results illustrate that, during the pandemic, people are turning to social media to share memories and reminisce.

Table 1. Example Quotes for Each Emotion

Emotion	Sample Quotes
Nostalgia	*As now, we are living during the COVID-19, restaurants are closed, and I cannot wait to go back to restaurant with some friends or family to restaurant again (P11).* *The joke of the post makes me laugh and it makes me happy to share it with another close people to remind us some good memories (P6).*
Enthusiasm	*We are both excited by this soon coming out so I wanted to share that with her (L10).* *I liked this video because this media often post some interesting cultural videos. I know the founder of the media and I like his work. By liking this video, I can show my support (P16).*

	When I post something, I want it to be something that shows what I like to do (L8).
Anger	*I just wanted to support this reporter who fights against natural disaster or unfair governmental decisions (P16).* *I wanted to raise awareness about Kosovo's Political Regulations (P20).* *It talks about the Situation of students in this pandemic and how it affected them and they are being ignored by the governments (P22).*
Apathy	*I try rather to remain detached from all this because I consider that it does not bring me anything and I think that on the contrary I find it a little puerile and boring to be dependent of social media. (L3)*

When nostalgia promoted participants to post on social media to relieve distress, they released information about their personal history and locational movements (i.e., previous, and current location information). As one participant stated, "Most of these posts are pictures taken during the vacations with my friends. I see Instagram more like a memory box" (P26). The privacy challenge with such information disclosures on social media is that, as the participant states, social media acts like a "memory box" and actors who can access a person's social media can view the locational patterns of the person, their friends, and details of their personal history that could be used for malicious purposes. For example, such information could inform a targeted phishing attack.

Enthusiasm and Personal Interests

Social validation motives have been found to be a primary motive driving public social media posts (Bazarova & Choi, 2014). Drawing from self-presentation theory (Baumeister, 1982), Omarzu (2000) suggests that seeking social approval is the default goal for most people and that to gain social approval, individuals will disclose information that presents their preferred, socially acceptable image to others. Posting on social media about interests, hobbies, and recreational pastimes helps users project a desired image to others.

We found that participants posted information about their interests when they were excited, enthusiastic, or inspired. A participant who frequently posted on Instagram stated, "I am more likely to go on social networks and comment on certain publications when I am in a good mood and enthusiastic" (L8). Self-presentation is generally considered to be motivated by a desire to make a favorable impression on others to gain social approval (Herring & Kapidzic, 2015). Our participants posted about their interests on social media to share their lives with others and present themselves in an appealing manner to their peers. For example, one participant stated: "I want people to see that I do interesting things" (L10). Another participant stated: "I post to share what I want people following me to see of my life" (L1). Our results indicate that enthusiasm prompts users to disclose information about their interests to obtain social approval. In fact, enthusiasm may result in users revealing more than they otherwise might, as one participant stated: "I think I may share extra information when I'm enthusiastic, as I forget about the danger of giving information on the internet" (P3).

Our results revealed that enthusiastic users are likely to share their interests and hobbies with their social media followers or friends in the hopes of being rewarded with likes. For example, one participant stated: "I noticed that when I got a lot of reactions (until 100 likes on a comment once), I tend to share and publish more. I believe this will of making people laugh was partially motivated by being accepted and recognized as a funny person. Having no reaction at all would have stopped me directly" (P12). The likes they receive on a post release a pleasurable emotional reaction of happiness in response to the fulfillment of a goal (i.e., to obtain social validation).

Anger and Political Affiliation

Social media has "revolutionized the way we experience politics, by engaging more citizens in the political process and enabling minority voices to be heard" (European Union, 2020, p. 1). Although viral political content often generates more attention, social media is also used as a place where people express their personal opinions about politics (Edgerly et al., 2016). We found that anger about governance, whether it be opinions about how a government is handling the pandemic or reactions to a political movement, prompted people to post on social media.

Self-clarification or identity clarification involves "talking about one's beliefs and opinions to clarify one's own position" (Prager et al., 1989, p. 564). We found that participants posted about their beliefs or opinions on political topics because they were angry. One participant illustrated their anger "I don't feel like we are making progress and we are living in the same worst conditions as we were before the revolution and I had to post it" (P22). This participant posted a picture to social media that depicts a group of Tunisian revolutionaries with a sign that states "Freedom For Tunisians." We found that our participants were motivated to talk about their opinions and beliefs online when they were angry about political decisions and movements.

When participants revealed their opinions because they were angry, they often revealed information about their political beliefs and affiliations. Posting political opinions clarifies a person's stance and can confirm their identity (e.g., as an anti-vaxxer or a member of the democratic party). For example, one participant was angered by online schooling and vented their frustration to their peers to be a part of a unified group against e-learning. This participant stated: "I wanted to share this idea and to spread it among my followers as most of them are students and they are living the same frustration as shown in the video. And by then why not reaching some government staff and must trigger their intentions to make some changes and care more about their youth" (P22). The participants indicated that they were motivated to represent the group they believe to be afflicted by the political landscape.

Our respondents were motivated to talk about their political beliefs and opinions when they were angry to confirm their identity and position. However, disclosing such information on social media confirms the identity to anyone who can access that person's social media. This could have unintended consequences, for example, research has shown that political affiliation determined from social media screening can influence hireability decisions (Roth et al., 2020).

IMPLICATIONS FOR RESEARCH AND PRACTICE

Drawing on the functional model of self-disclosure and the disclosure decision model, we study how emotion gives salience to motives that result in information disclosures on social media. Our study contributes to the privacy literature by showing how emotion is associated with motives from the functional model of self-disclosure and what type of information disclosures result. We highlight the importance of emotion in the privacy decision making context.

Users have a difficult time understanding their social media audience and that includes underestimating the size of that audience (Bazarova & Choi, 2014; Bernstein et al., 2013).

Herring and Kapidzic (2015) explain that privacy settings give one a measure of control over one's audience. However, many social media users have difficulty managing privacy controls (Madden, 2012). Young adults are especially prone to being "trigger-happy" and posting social media content that they later regret sharing (Madden, 2012). Our results show that emotional social media posting does occur. Advances in technology may make it harder for people to determine the future consequences of information disclosures (Laufer & Wolfe, 1977). People may regret social media information disclosures once the consequences are realized.

Social media organizations may be able to stop private information from being posted on their platform by developing algorithms that are able to detect the type of information to be revealed before it is posted and encouraging users to carefully consider if they want to release it. Twitter recently launched a feature to detect mean tweets (Diaz, 2021). When a mean tweet is detected, the feature will prompt the user with "Want to review this before tweeting" to encourage users to think twice before posting. Similar features could be developed to detect emotional or privacy compromising posts and provide appropriate guidance. Another option would be user training or social media campaigns to increase awareness of data and social media privacy. Space usually reserved for ads could be replaced by information and warnings to users browsing the platform to stay proactive and conscious about their privacy. Social media privacy could also be a consideration for new privacy regulations.

LIMITATIONS AND FUTURE RESEARCH

Our study is not without limitations. Our data was collected in Europe which has implemented wide-ranging privacy regulations through the recently enacted General Data Protection Regulation. Because of this visible regulation, Europeans may be more privacy aware than people in other countries. Future research should examine emotional privacy decisions in other cultural contexts. Future studies may also want to consider the role of privacy awareness generated by governmental regulation on how people manage their social media privacy.

CONCLUSION

Our results show that emotions can trigger people to disclose personal information on social media. Our data reveals that users feeling nostalgic for pre-pandemic life post to social media to relieve distress by reminiscing about past happier times. Users post their previous experiences and vacations to cope with the burden of the pandemic, but in doing so, release location and personal history information. We find that enthusiasm is associated with people posting about their interests, hobbies, and recreational pastimes on social media to gain social validation. Enthusiastic users are excited to share their interests and hobbies with their followers in hopes that they rewarded with likes. Anger is associated with posts that reveal the political viewpoints of users intended to confirm their political identities and positions. Overall, our findings show that users make emotionally driven information disclosures on social media for several different reasons.

REFERENCES

Bagozzi, R. P., Gopinath, M., and Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science, 27*(2), 184-206.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91*(1), 3-26.

Bazarova, N. N. and Choi, Y. H. (2014). Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication, 64*(4), 635-657.

Bazarova, N. N., Choi, Y. H., Schwanda Sosik, V., Cosley, D., and Whitlock, J. (2015). *Social sharing of emotions on Facebook: Channel differences, satisfaction, and replies.* Paper presented at the Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, Vancover, Canada.

Bernstein, M. S., Bakshy, E., Burke, M., and Karrer, B. (2013). *Quantifying the invisible audience in social networks.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems, Paris, France.

Derlega, V. J. and Grzelak, J. (1979). Appropriatness of self-disclosure. In G. J. Chelune (Ed.), *Self-Disclosure: Origins, Patterns, and Implications of Openness in Interpersonal Relationships* (pp. 151-176). San Francisco, CA: Jossey-Bass.

Diaz, J. (2021). Want to Send a Mean Tweet? Twitter's New Feature Wants You To Think Again Retrieved May 9, 2021, from https://www.npr.org/2021/05/06/994138707/want-to-send-a-mean-tweet-twitters-new-feature-wants-you-to-think-again

Edgerly, S., Thorson, K., Bighash, L., and Hannah, M. (2016). Posting about politics: Media as resources for political expression on Facebook. *Journal of Information Technology & Politics, 13*(2), 108-125.

European Union (2020). Social media influences our political behaviour and puts pressure on our democracies, new report finds Retrieved May 4th, 2021, from https://ec.europa.eu/jrc/en/news/social-media-influences-our-political-behaviour-and-puts-pressure-our-democracies-new-report-finds

Gammon, S. and Ramshaw, G. (2020). Distancing from the present: Nostalgia and leisure in lockdown. *Leisure Sciences, 43*, 1-7.

Harber, K. D. and Cohen, D. J. (2005). The emotional broadcaster theory of social sharing. *Journal of Language and Social Psychology, 24*(4), 382-400.

Hemenover, S. H. (2003). The good, the bad, and the healthy: Impacts of emotional disclosure of trauma on resilient self-concept and psychological distress. *Personality and Social Psychology Bulletin, 29*(10), 1236-1244.

Herring, S. C. and Kapidzic, S. (2015). Teens, Gender, and Self-Presentation in Social Media. In J. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences, Second Edition* (pp. 146-152). Oxford, UK: Elsevier.

Laufer, R. S. and Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues, 33*(3), 22-42.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist, 46*(8), 819-834.

Madden, M. (2012). Privacy management on social media sites Retrieved May 4, 2021, from https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_Privacy_management_on_social_media_sites_022412.pdf

Molla, R. (2021). Posting less, posting more, and tired of it all: How the pandemic has changed social media Retrieved May 4, 2021, from https://www.vox.com/recode/22295131/social-media-use-pandemic-covid-19-instagram-tiktok

Myrick, J. G. (2015). Emotion regulation, procrastination, and watching cat videos online: Who watches Internet cats, why, and to what effect? *Computers in Human Behavior, 52*, 168-176.

Omarzu, J. (2000). A disclosure decision model: Determining how and when individuals will self-disclose. *Personality and Social Psychology Review, 4*(2), 174-185.

Prager, K. J., Fuller, D., and Gonzalez, A. (1989). The Function of Self-Disclosure. *Journal of Social Behavior and Personality, 4*(5), 563-580.
Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review, 1*(1), 60-85.
Roth, P. L., Thatcher, J. B., Bobko, P., Matthews, K. D., Ellingson, J. E., and Goldberg, C. B. (2020). Political affiliation and employment screening decisions: The role of similarity and identification processes. *Journal of Applied Psychology, 105*(5), 472.
Seattle Times Staff (2020). What we miss from the time before the coronavirus — and what we hope never comes back Retrieved May 4, 2021, from https://www.seattletimes.com/life/heres-what-we-miss-from-the-time-before-the-coronavirus-and-what-we-hope-never-comes-back/
Strauss, A. and Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications.
van Ingen, E., Utz, S., and Toepoel, V. (2016). Online coping after negative life events: Measurement, prevalence, and relation with internet activities and well-being. *Social Science Computer Review, 34*(5), 511-529.
Vermeulen, A., Vandebosch, H., and Heirman, W. (2018). # Smiling,# venting, or both? Adolescents' social sharing of emotions on social media. *Computers in Human Behavior, 84*, 211-219.

DECISION SCIENCES INSTITUTE

How National Culture Influence the Covid-19 Pandemic

Yundong Huang
St. Edward's University
Email: yhuang2@stedwards.edu

ABSTRACT

National culture reflects the collective behaviors and thinking patterns of the residences in a country. It potentially affects how people react to a Pandemic and local government's public health policies, such as a lockdown order. This quantitative study examines the moderating effect of national culture on the relationship between government stringency and COVID-19 virus infection. The structural equation modeling (SEM) technique is used to discover the influence of national culture, measured by Hofstede's Six-cultural-dimensions. The results show that Individualism (IDV) has a significant moderating effect, while Power Distance (PDI), Masculinity (MAS), and Long-term orientation (LTO) have marginal moderating effects. The findings indicate high individualism culture negatively moderate (weaken) the effectiveness of government stringency for containing COVID-19 pandemic.

KEYWORDS: National Culture, Covid-19, Cultural Dimension

INTRODUCTION

In December 2019, a new type of Corona Virus was identified and quickly spread around the world. It was later named as severe acute respiratory syndrome coronavirus 2 (SARS‑CoV‑2) (CDC Covid-19 Response Team, 2020). It developed as a "Public Health Emergency of International Concern" in January 2020, and became an ongoing Pandemic on March 2020, named COVID-19 (WHO, 2020a). As of July 30, 2020, approximately 17 million people infected this virus in 188 countries, resulting in 668,000 deaths (CSSE, 2020). The virus is primarily spread by close contact between human (WHO, 2020b; CDC, 2020)

When the scientists are still working on vaccine and treatment for the COVID-19, the major way of government's response to the pandemic is to reduce the circumstance of close contact between people. Many emergency public health orders are executed, such as lockdown, shelter in place, mandatory social distance, etc. (Hale et al, 2020). It's still an ongoing debate on what is the proper level of lockdown according to the infection numbers since the lockdown will severely affect economic development.

According to the Oxford Coronavirus Government Response Tracker (OxCGRT) published by Oxford University, almost every country has implemented some sort of government control over the resident's traveling & gathering activities (2020). The effectiveness of these government controls (stringency level) varies among different countries. Therefore, a direct question is how well these government orders are strictly followed by the public.

It is a question very hard to answer since it involves massive individual-behavioral observation which is practically impossible to conduct at the current stage. You may see some news about people or businesses get punished for violating public health orders. However, it's hard to say how many others are violating the orders. There are no statistics on this critical issue, but we can observe it indirectly from the increasing numbers of COVID-19 infection. Countries with similar government stringency levels may have a very different situation on infection number, infection rate, growth rate, etc. There are multiple explanations for these variances, such as income level, economic development level, medical infrastructure level, urbanization level, population density, and national culture.

Even limited evidence, it's already a widely discussed topic that national culture may affect people's reactions to the COVID-19 pandemic (Frey et al. 2020). In countries with collective cultures, such as Japan, the public may follow the government orders more faithfully. While for countries with individualism cultures, such as the United States, people may follow their own judgments on social behavior, which may potentially lead to a faster spread of the virus. However, these speculations need to be examined carefully.

This study is to provide a rigorous statistical analysis of whether or not national culture affects the spread of COVID-19. Specifically, this study proposes that national culture moderates the effectiveness of government control on COVID-19 pandemic. In other words, people with certain cultural backgrounds are more likely to cooperate with government orders that mitigate the impact of COVID-19.

Scholars have noticed the potential influence of national culture on the pandemic. For example Biddlestone, Green, and Douglas (2020) found national culture affected people's intention on keeping social distance and hygiene behaviors. Jiang, Wei, and Zhang (2020) discovered individualism was significantly related with COVID-19 case increase rate and death rate.
However, national culture shouldn't be a determinant for any virus infection. It's a scientific fact that people get infected only because the virus enters their body when they have close contact with the source of the virus (CDC, 2020). Other factors like national culture and urbanization level may influence the chance of close contact between people, but not a direct reason for virus infection. Therefore this study will examine the moderating effects of national culture.

NATIONAL CULTURE

National culture has been a very important research stream in social science (Hofstede, 2001). The basic argument of culture study is that one’s cultural background may potentially affect his/her opinions and actions. When conducting research on culture, the two most important questions are: “What is culture?” and “How to measure it?”

This study adopts Hostede’s widely cited definition: “Culture is the collective programming of the mind which distinguishes the members of one category of people from another” (Hofstede, 1984). However, one definition was far from enough. Researchers were unclear about what was included in the term “culture”.

Geert Hofstede conducted a pioneering study on national culture between 1967 and 1973 which was called cultural dimension theory. This theory is one of the most cited pieces of research in the field of cross-culture comparative management. His original study emerged from two in-house attitude questionnaire surveys in subsidiaries of IBM (Hofstede, 1984).

Base on the survey, Hofstede proposed four cultural dimensions, each placed on a continuum ranging scores from high to low. A set of scores for each dimension represent a nation's culture. The dimensions are power distance, uncertainty avoidance, individualism/ collectivism, and masculinity/femininity (Hofstede, 2011).

Together with Micheal Bond, Hofstede later identified a fifth dimension which was first termed 'Confucian dynamism' and then renamed as 'time orientation'. 'Time orientation' is argued to measure two contrasting poles that distinguish short-term orientated cultures from the long-term orientated ones. In 2010, the sixth dimension, indulgence versus restraint, was added to describe the attitude toward pursuing individual happiness. It’s based on the findings from the World Value Survey by Misho Minkov (Minkov, 2009; Hofstede, 2011). Hofstede’s cultural dimension theory provided a very useful quantitative measurement for studies on national culture and became one of the most cited works on cross-cultural management studies. The latest database from Hofstede Insight includes 111 countries based on the survey results in 2015.

The following table 1 shows the 6 dimensions of national culture:

Table 1: Six Dimensions of National Culture

Cultural Dimension	Content
Individualism VS collectivism (IDV)	The high side of this dimension, called Individualism, can be defined as a preference for a loosely-knit social framework in which individuals are expected to take care of themselves and their immediate families only. Its opposite, Collectivism, represents a preference for a tightly-knit framework in society in which individuals can expect their relatives or members of a particular in-group to look after them in

	exchange for unquestioning loyalty.
Power Distance (PDI)	The degree to which the less powerful members of a society accept and expect that power is distributed unequally. The fundamental issue here is how a society handles inequalities among people. People in societies exhibiting a large degree of power distance accept a hierarchical order in which everybody has a place and which needs no further justification. In societies with low power distance, people strive to equalize the distribution of power and demand justification for inequalities of power.
Masculinity VS femininity (MAS)	The masculinity side of this dimension represents a preference in society for achievement, heroism, assertiveness, and material reward for success. Society at large is more competitive. Its opposite, femininity, stands for a preference for cooperation, modesty, caring for the weak, and quality of life. Society at large is more consensus-oriented.
Uncertainty avoidance (UAI)	The uncertainty avoidance dimension expresses the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity. The fundamental issue here is how a society deals with the fact that the future can never be known: should we try to control the future or just let it happen? Countries exhibiting strong UAI maintain rigid codes of belief and behavior and are intolerant of unorthodox behavior and ideas. Weak UAI societies maintain a more relaxed attitude in which practice counts more than principles.
Long-term VS short-term orientation (LTO)	Societies with a short-term orientation generally have a strong concern with establishing the absolute Truth. They are normative in their thinking. They exhibit great respect for traditions, a relatively small propensity to save for the future, and a focus on achieving quick results. In societies with a long-term orientation, people believe that truth depends very much on situation, context and time. They show an ability to adapt traditions to changed conditions, a strong propensity to save and invest thriftiness, and perseverance in achieving results.
Indulgence VS Restraint (IVR)	Indulgence stands for a society that allows relatively free gratification of basic and natural human drives related to enjoying life and having fun. Restraint stands for a society that suppresses gratification of needs and regulates it by means of strict social norms.

Source: Hofstede Insight (hofstede-insights.com)

HYPOTHESIS

The six dimensions of national culture reflect people’s collective values and behavior patterns in a society. Each of them may potentially affect how people follow their local government’s orders on COVID-19 pandemic. National culture does not directly cause or prevent virus infection, but it has an influence on people's attitudes and reactions to the pandemic. For example, Biddlestone et al. (2020) found that national culture affects people’s intention on keeping social distance, which is one of the major methods governments recommended or request to the public.

Individualism VS collectivism (IDV) refers to how tight individuals are connected to each other in a society. In high Individualism countries, like the United States, an individual's choice and decision are respected; people are expected to take care of themselves rather than their group; personal freedom is preferred over social obligation. In the circumstance of COVID-19 pandemic, people may regard their personal needs as more important than the public interest. Therefore, people from individualism culture will less likely to follow government control.

On the other side, in collectivism cultures, like Japan, people prefer social obligation over personal interest; individuals are more closely connected to each other; harmony and loyalty are emphasized within social groups. In COVID-19 pandemic, people are more likely to sacrifice personal freedom for public safety. Previous studies also showed that peer-pressure is stronger in collectivism cultures to force people to follow the social norms (Chen et al. 1998). Therefore, the first hypothesis should be:

Hypothesis 1: Individualism culture negatively moderates the effectiveness of the government's control of COVID-19 pandemic.

Power Distance (PDI) refers to the acceptance of the unequal social power distribution. In high power distance culture, like Mexico, people are more likely to accept social hierarchies; orders from higher-level are more likely to be followed rather than challenged. Reversely, low power distance countries, like Denmark, people emphasize on equality and justice. Therefore, government orders on the COVID-19 pandemic may be obeyed well in high power distance countries.
Government restrictions impacted the public in different ways. Large corporations, small businesses, a variety of professions, income levels, all these differences may affect people's attitudes on the fairness of the government. If certain groups of people in society regard the government order is unfair, how will they react to it? People with high power distance cultural backgrounds may complain but take it anyway. However, people from low power distance cultures may strive for justice, such as compensations. This lead to our second hypothesis:

Hypothesis 2: High power distance culture positively moderates the effectiveness of the government's control of COVID-19 pandemic.

Masculinity (MAS) refers to cultures that encourage heroism and assertiveness. In this type of culture, people eager for achievements and material rewards. The other side of this dimension is called Femininity, which prefers cooperation, modesty, caring for the weak, and quality of life. This cultural dimension generally describes the society is competitive or cooperative. The effectiveness of government restrictions requires the public to cooperate. Therefore, the third hypothesis should be:

Hypothesis 3: Masculinity culture negatively moderates the effectiveness of the government's control of COVID-19 pandemic.

Uncertainty avoidance (UAI) describes how people deal with the unpredictable future. People from high uncertainty avoidance culture will make great efforts to control the future, and feel uncomfortable with uncertainty. Low uncertainty avoidance cultures have a more lenient opinion to the future.

The future of COVID-19 pandemic is hard to predict now. At the current stage, very little information is known about this new type of virus. There is no available vaccine or effective treatment for it. As a type of contagion, the infection of the virus is also due to probability. Also, the severity of the symptom is case by case. Some infected patients have weak symptoms, even no symptoms; others have life-threatening symptoms (such as system inflammatory response syndrome, SIRS). COVID-19 caused a high level of uncertainty to both society and individuals. When people know joining a party has a chance to get infected, a chance to get fined, and a chance to die or recover, how will people react to it? People from high UAI cultures are more likely to reduce the risk of COVID-19 infection. This lead to the fourth hypothesis of this study:

Hypothesis 4: Uncertainty avoidance positively moderates the effectiveness of the government's control of COVID-19 pandemic.

Long-term orientation (LTO) reflects the culture that is more pragmatic. People from LTO cultures believe that truth depends on the situation, rather than traditions. The other side of this dimension is called short-term orientation, in which people "prefer to maintain traditions and norms while viewing societal change with suspicion" (Hostede Insight, 2020). During a pandemic, whether or not to maintain the social norm is critical for the effectiveness of the government order. Actually, the orders like shelter-in-place are a very traditional government reaction for disasters. This lead to our fifth hypothesis:

Hypothesis 5: Long-term orientation negatively moderates the effectiveness of the government's control of COVID-19 pandemic.

Indulgence VS Restraint (IVR) refers to the attitude of the society on pursuing personal enjoyment that is related to natural human desires. Indulgence culture, like Puerto Rico, let people freely chasing fun and enjoyment of life. Restraint culture, like Egypt, has strong social norms that suppress these needs. In the COVID-19 pandemic, many government controls have a great impact on people's enjoyment of life, for example, the closure of movie theaters, sports events, bars and restaurants, public parks. How will people react to it? It may be harder for people from indulgence culture to suppress their needs for entertainment. For people from restraint culture, the temporary extra control from the government may not be a significant change. This lead to our sixth hypothesis:

Hypothesis 6: Indulgence culture negatively moderates the effectiveness of the government's control of COVID-19 pandemic.

The influence of national culture on people's behavior during a pandemic is a very new topic in the literature of cross-cultural studies. There are very few previous studies and/or established theories that can be used to support our hypotheses. This study is an exploratory study. The focal hypothesis is our first hypothesis that examines the role of individualism and collectivism.
In general, there are very few researches discuss the role of national culture on a pandemic. Therefore, most of the hypotheses in this study are explorative. More theoretical discussions and empirical evidence are needed in the future.

METHODOLOGY

The dependent variable in this study is the accumulated cases of COVID-19 infection in a country. The source of data is World Health Organization (WHO) retrieved from (covid19.who.int) in July 2020. Accumulated case numbers may be inaccurate due to the development level of the health system, including test capability and reporting system. Furthermore, data for other variables in this study, such as culture dimension, are also limited in many countries in the world. Therefore, this study only used the data of the Organization for Economic Co-operation and
Development (OECD) member countries. The thirty-seven member countries of OECD can be found at (www.oecd.org/about/members-and-partners). WHO data were retrieved from the first day a country reports its first COVID-19 case to July 9th, 2020.

The determinant of accumulated COVID-19 cases number is only due to the virus infection, which is mainly due to the close contact between people. Since tracking every individual's behavior in every country is impossible, most research institutions use computer simulation to evaluate the infection number. This research adopted the simulation model developed by Martin Eichner (Epimos GmbH) and Markus Schwehm (ExploSYS GmbH) supported by University of Tübingen and IMAAC NEXT Association. The equations of simulation can be found at http://version-1.1.covidsim.eu/assets/Model_description_Version_1.1.pdf

The simulation data shows the theoretical infection case numbers if no action is taken by the government and/or individuals. The simulation case number only related to a country's population and the duration of the pandemic. Two more control variables are added to control the potential impact of urbanization and income level. Urbanization refers to the percentage of the population live in urban-area. It better described how close people live together in a country than population density, which refers to the number of people lived per square kilometer.

Countries like Canada and Australia have low population density due to their large territory, but most of their population lived in high density in urban-area. Income level is measure by GDP per capita, which may affect people's financial capability on seeking medical treatment for the virus. Both control variable data are retrieved from World Bank Database (data.worldbank.org) in July 2020.

Government restriction on COVID-19 spread is measure by The Oxford COVID-19 Government Response Tracker (OxCGRT). This study used one of the indexes from OxCGRT, called government stringency index, which reflects the level of government restriction on various aspects of orders, including containment and closure policies; economic policies; health system policies; miscellaneous policies. The detail method of OxCGRT can be found at https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/index_methodology.md

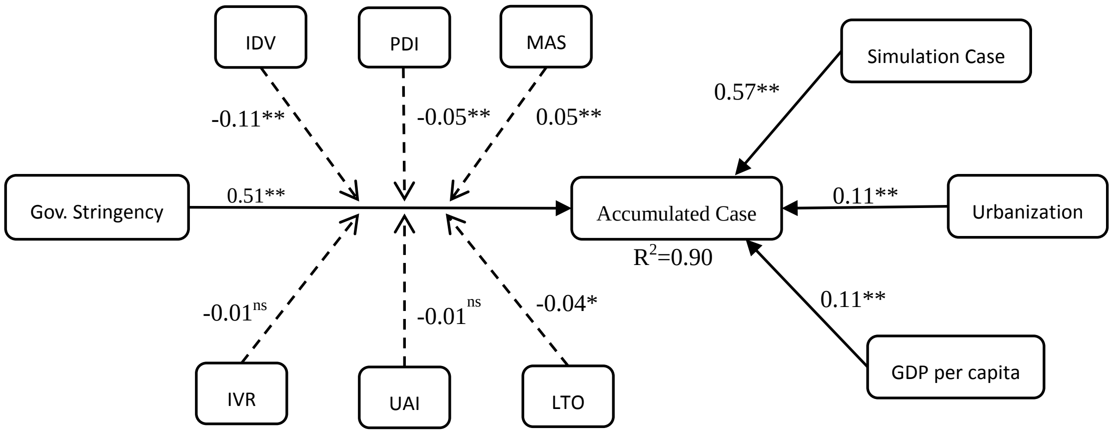
Culture dimension variables data were retrieved from geert-hofstede.com on July 2020. These data include all six dimensions introduced previously in the 37 OECD countries. These cultural dimensions are survey-based scores from 0 to 100 published in December 2015. The only missing data is the IVR score for Israel. The data set was downloaded from https://geerthofstede.com/research-and-vsm/dimension-data-matrix.

This study applied the structural equation modeling (SEM) technique to analyze the data. A total of 5305 observations are included. The software used in this analysis is WarpPLS 7.0, a variance-based Partial Least Squares (PLS) analysis tool.

RESULTS

The results of SEM analysis are presented in Figure 1. The numbers beside the arrows are the (standardized) path coefficients indicating the strengths of the multivariate associations between variables. The statistical significance level of the path coefficients are: ** refers to $p \leqslant 0.001$, * refers to $p \leqslant 0.005$, and ns refers to non-significant. In Figure 1, solid arrows represent the direct association between two variables; dashed arrows represent the moderating effect of a variable on a correlation. R-squared coefficient is also presented in Figure 1, which shows the percentage of variance explained by the variables (including control variables) that point to the endogenous variables in the research model.

Figure 1 Research Model and Path Coefficient:



** (p≤0.001), * (p≤0.005), ns (non-significant)

Overall, this research model explained 90% (R2=0.90) of the variance of accumulated COVID-19 cases in our sample countries. Simulation case numbers are the theoretical estimate based on a country's population and duration of the pandemic. The SEM result shows it has a very high path coefficient and high effect sizes (beta=0.57**, ES=0.48). The effect sizes are calculated as the absolute values of the contribution to the R-squared coefficient. Government stringency is also significantly associated with accumulated COVID-19 cases, with high effect sizes (beta=0.51**, ES= 0.39). These two independent variables contribute 87% of the R-squared coefficient.

The two control variables are also significantly associated with the dependent variable. Urbanization level positively associated with COVID-19 cases (beta=0.11**, ES=0.01). GDP per capita also has positive contribution (beta=0.11**, ES< 0.01). However, the low effect sizes of the two control variables indicate a very weak influence on the dependent variable. For moderating effects of national culture, SEM analysis shows mixed results. Individualism (IDV) negatively moderates the relation between government stringency and COVID-19 cases (beta=-0.11*, ES= 0.03). Therefore, hypothesis 1 is supported. Hypothesis 1 is the focal hypothesis in this study. It indicates national culture with high individualism will weaken the effectiveness of government control on the COVID-19. On the other side, the collectivism culture will strengthen it.

Power distance (PDI) negatively moderates the relation (beta=-0.05**, ES=0.02), which is hypothesized as a positive effect. So, hypothesis 2 is rejected. Masculinity (MAS) positively moderates the relation (beta=0.05**, ES=0.02), which is also reversed as hypothesis 3. Uncertainty avoidance (UAI) show non-significant path coefficient (beta=-0.01ns, ES<0.01). Therefore hypothesis 4 is rejected. Long-term orientation (LTO) shows negative moderating effects (beta=-0.04*, ES= 0.02).Therefore, hypothesis 5 is supported. Indulgence VS Restraint (IVR) is hypothesized as having a negative moderating effect. However, it shows a non-significant coefficient (beta=-0.01ns, ES<0.01). Thus, hypothesis 6 is rejected. Table 2 summarized the path coefficients and effect sizes of all variables:

Table 2 Path Coefficient and Effect Sizes:

Variables	***Path Coefficient***	***Effect Sizes***
Gov. Stringency	0.51, p≤0.001	0.39
Simulation Case	0.57, p≤0.001	0.48
Urbanization	0.11, p≤0.001	0.01
GDP per capita	0.11, p≤0.001	<0.01
IDV* Gov. Stringency	-0.11, p≤0.001	0.03
PDI* Gov. Stringency	-0.05, p≤0.001	0.02
MAS* Gov. Stringency	0.05, p≤0.001	0.02
UAI* Gov. Stringency	-0.01, p=0.35	<0.01
LTO* Gov. Stringency	-0.04, p=0.002	0.02
IVR* Gov. Stringency	-0.01, p=0.23	<0.01

Table 2 shows the variance of accumulated cases of COVID-19 is mainly explained by the theoretical estimation (Simulation Case) based on population and duration, and the government restrictions (Gov. Stringency). In addition, national culture contribute approximately 9% of the explanatory power through the moderating effects. Another model without national culture variables is also analyzed. It confirms that without national culture the R-square of the dependent variable dropped from 0.90 to 0.81. (See appendix 1: SEM Results without Moderator).

Table 3 shows the correlation coefficients and their significant level (p value) among all variables. As shown on table 2, many of the correlations are statistically significant, which indicate potential collinearity issue. Therefore, a full collinearity check is conducted by calculating the variables variance inflation factor (VIF) score. Table 4 presents the VIF score of the variables. It shows that all VIF scores are lower than the 3.3 criteria, which indicate the colinearity level in this assessment model is acceptable.

Huang How National Culture Influence the Covid-19 Pandemic

Table 3 Correlation Matrix of All Variables

	Accu. Case	Sim. Case	Gov. String.	IDV	PDI	MAS	UAI	IVR	LTO	Urban	GDP Capi	IDV* Str	PDI* Str	MAS* Str	IVR* Str	UAI* Str	LTO* Str
Accu. Case	1.00																
Sim. Case	0.81**	1.00															
Gov. String.	0.68**	0.34**	1.00														
IDV	0.01	0.05**	-0.12**	1.00													
PDI	0.03	0.00	0.11**	-0.51**	1.00												
MAS	0.06**	0.06**	0.06**	0.05*	0.27**	1.00											
UAI	0.03	0.00	0.13**	-0.63**	0.57**	0.20**	1.00										
IVR	0.09**	0.02	-0.01	0.12**	-0.25**	-0.01	-0.38**	1.00									
LTO	-0.06**	0.04	-0.13**	-0.04*	0.16**	0.13**	0.21**	-0.62**	1.00								
Urban	0.01	0.06**	-0.14**	0.20**	-0.29**	-0.22**	-0.12**	0.48**	-0.08**	1.00							
GDPCapi	0.02	0.01	-0.14**	0.49**	-0.59**	-0.10	-0.51**	0.36**	-0.06**	0.37**	1.00						
IDV*Str	0.12**	0.12**	0.05**	0.14**	-0.08**	0.04	-0.03**	-0.17**	0.24**	-0.01	0.12**	1.00					
PDI*Str	-0.01	-0.02	-0.01	-0.08**	0.17**	-0.03**	-0.01**	0.08**	-0.09**	-0.05**	-0.12**	-0.53**	1.00				
MAS*Str	0.01	-0.01	0.04*	0.05	-0.03	-0.15**	-0.19**	0.10**	-0.14**	-0.08**	0.02	0.05**	0.18**	1.00			
IVR*Str	0.17**	0.15**	0.17**	-0.18**	0.07**	0.09**	0.10**	0.06**	0.00	0.04	-0.07**	0.09**	-0.15**	0.04*	1.00		
UAI*Str	-0.05**	-0.08**	-0.04*	-0.03	-0.01	-0.18**	-0.01	0.10**	-0.16**	0.02	-0.08**	-0.64**	0.62**	0.17**	-0.39**	1.00	
LTO*Str	-0.10**	-0.09**	-0.10**	0.25	-0.09**	-0.13**	-0.16**	0.00	-0.04	-0.02	0.12**	-0.01	0.13**	0.05**	-0.66**	0.26**	1.00

** (p≤0.001); *(p≤0.005)

Table 4 Variance Inflation Factor (VIF) Score

	Sim. Case	Gov. String.	Urban	GDPCapi	IDV*Str	PDI*Str	MAS*Str	IVR*Str	UAI*Str	LTO*Str
VIF Score	1.77	2.79	1.285	1.388	2.236	1.659	1.824	2.084	1.868	2.486

CONCLUSION

The purpose of this study is to explore how national culture affects COVID-19 pandemic. As discussed previously, national culture doesn't directly cause or prevent people from virus infection. However, it potentially affects people's attitudes and reactions to the government's orders which are designed to reduce the close contact between people. In other words, national culture changes the effectiveness of the government control's stringency level, through which it influences the COVID-19 affection cases. The six dimensions of national culture are examined individually. Four of them influence the pandemic in different ways.

The SEM analysis of this study confirmed that national culture is an important moderator on the relationship between government stringency and the accumulated case of COVID-19. It overall contributes about 9% of the explanatory power. Specifically, individualism (IDV) will weaken the effectiveness of government stringency. As hypothesized, people from individualism cultures are more likely to place their personal decision and interest over others. Therefore, they are less likely to strictly follow the orders from the government regarding the COVID-19. This empirical finding may explain some countries faster contain the virus worldwide, such as Japan and South Korea, which are collectivism (low-individualism) countries.

Power distances (PDI) also weaken the relationship between government stringency and the accumulated case of COVID-19. It means that a society that has a high tolerance for the unequal distribution of social power will have a weaker performance on COVID-19 control. This may partially explain why Belgium (PDI = 65) contains the virus slower than its neighbor country Germany (PDI =35), with similar government stringency levels. One thing that should be noted that this moderating effect is very weak (ES=0.016).

Masculinity (MAS) culture improves the effectiveness of COVID-19 control. It indicates a country with competitive culture is more effective than the cooperative culture when fighting with the virus. This finding is a little counterintuitive. A possible explanation is that the cooperative culture (feminism) focuses more on "quality of life" which is harmed by government stringency on COVID-19. This difference can be observed from the comparison between Sweden (MAS=5) and Switzerland (MAS=70). The former has 8225 CVOID-19 cases per 1 million populations; the latter only has 4260 cases per 1 million populations (Worldometers, 2020).

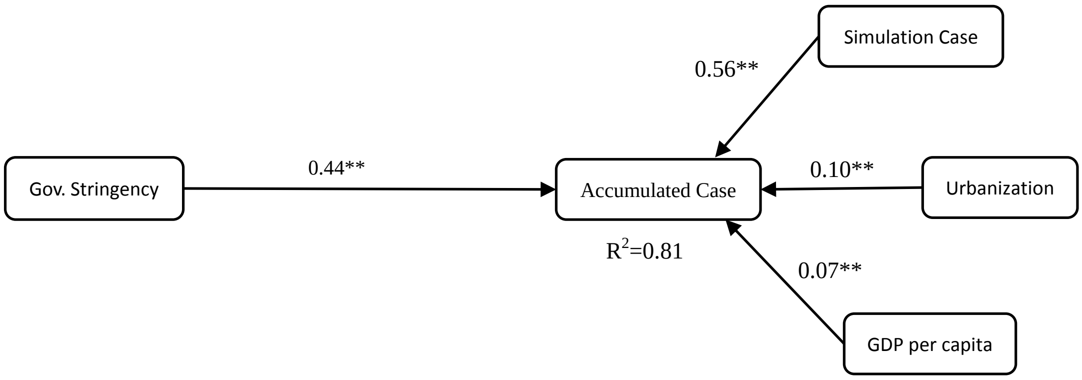
Long-term orientation (LTO) weakens the effectiveness of COVID-19 control. As hypothesized, it is mainly due to whether or not the social norm and traditions should be strictly followed. However, the SEM analysis only shows marginal support. The path coefficient is marginally significant (p=0.002); the effect sizes is low (ES=0.019). Therefore, to be conservative, it is hard to conclude that LTO has a practical influence on the pandemic.

Uncertainty avoidance (UAI) has no influence on the pandemic. Theoretically, people from high UAI cultures tend to avoid risky activities, such as traveling and gathering during the pandemic. This tendency should strengthen the government's control over the virus spread. However, the empirical analysis didn't support this prediction. Similarly, Indulgence VS Restraint (IVR) was hypothesized to influence the virus spread through the stringency of the social norm on pursuing personal happiness. However, the empirical analysis didn't support it either.

LIMITATION

This study is subject to some limitations. Firstly, government stringency data used in this study are national-level data. However, many countries have province and city level controls which were not consistent with the national level COVID-19 control. In this case, the national stringency level may not fully reflect all reactions of the local government. Secondly, National culture data are also roughly averages. Actually, many countries have very different regional cultures which may have a stronger influence on the pandemic than its national culture. Thirdly, there are also some omit factors have a potential influence on people's reaction of government control, such as the influence of media, influence of the religious group, and political situations. These limitations may be eliminated when more data is available. For example, OxCGRT project (2020) recently started to report the second-level government stringency index in the United States.

APPENDIX 1: SEM RESULTS WITHOUT MODERATOR



** (p≤0.001), * (p≤0.005), ns (non-significant)

REFERENCES

Biddlestone, M., Green, R., & Douglas, K. (2020). Cultural orientation, powerlessness, belief in conspiracy theories, and intentions to reduce the spread of COVID-19. ***British Journal of Social Psychology***. 59, 663-673.

CDC COVID-19 Response Team (2020). Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States, February 12–March 16, 2020. ***MMWR Morb Mortal Wkly Rep***, *69*(12), 343-346.

Centers for Disease Control and Prevention (2020). How COVID-19 Spreads. Retrieved July 09, 2020, from https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html

Center for Systems Science and Engineering (2020). COVID-19 Dashboard by the Johns Hopkins University (JHU). Retrieved July 07, 2020, from https://coronavirus.jhu.edu/data

Chen, C. C., Chen, X. P., & Meindl, J. R. (1998). How can cooperation be fostered? The cultural effects of individualism-collectivism. ***Academy of management review***, *23*(2), 285-304.

Frey, C. B., Chen, C., & Presidente, G. (2020). Democracy, Culture, and Contagion: Political Regimes and Countries Responsiveness to Covid-19. ***Covid Economics***, *18*, 1-20.

Hale, T., Petherick, A., Phillips, T., & Webster, S. (2020). Variation in government responses to COVID-19. ***Blavatnik school of government working paper***, *31*.

Hofstede, G. (1984). Cultural dimensions in management and planning. ***Asia Pacific journal of management***, *1*(2), 81-99.

Hofstede, G. (2001). ***Culture's consequences: Comparing values, behaviors, institutions and organizations across nations***. New York: Sage.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. ***Online readings in psychology and culture***, 2(1), 2307-0919.

Hofstede Insight (2020). NATIONAL CULTURE. Retrieved July 07, 2020, from https://hi.hofstede-insights.com/national-culture

Jiang, S., Wei, Q. & Zhang, L. (2020). Impacts of Cultural Difference on the Transmission of COVID-19：Individualism vs. Collectivism. 10.13140/RG.2.2.34274.96966.

Eichner, M. and Schwehm, M. (2020), Mathematical description of the CovidSIM model. Retrieved July 09, 2020 from http://covidsim.eu/assets/Model_description_Version_1.1.pdf

Minkov, M. (2009). Predictors of differences in subjective well-being across 97 nations. **Cross-Cultural Research**, 43(2), 152-179.

Oxford University (2020), Oxford Coronavirus Government Response Tracker (OxCGRT). Retrieved July 07, 2020, from https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

World Bank (2020). World Bank Database. Retrieved July 07, 2020, from https://data.worldbank.org/

World Health Organization (2020a). Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). Retrieved July 09, 2020, from https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).

World Health Organization (2020b). Q&A on coronaviruses (COVID-19). Retrieved July 09, 2020, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses

Worldometers (2020). Covid-19 Coronavirus Pandemic. Retrieved July 09, 2020, from https://www.worldometers.info/coronavirus/

DECISION SCIENCES INSTITUTE

A Preliminary Comparison of Instructor and Student Perspectives between Online and Face-to-Face Education During the Pandemic

Lynn A. Fish, Ph.D.
Canisius College
fishl@canisius.edu

Coral R. Snodgrass, Ph.D.
Canisius College
snodgras@canisius.edu

ABSTRACT

During the pandemic as part of an ongoing research stream, a comparison of instructors and business students' perspectives at an AACSB Jesuit, Catholic University with a strong focus on teaching face-to-face revealed that instructors and students still prefer face-to-face instruction on most individual and program factors. Given the impact the pandemic has on education, this research offers insight into future education to meet student perceptions.

KEYWORDS: Instructor Perspectives, Student Perspectives, Online, Face-to-Face

INTRODUCTION

While the University in this study offered online courses and programs, the face-to-face (FTF) educational methodology was the norm at the institution of study. In prior separate surveys for instructors (Fish & Snodgrass, 2018a, 2018b) and business students (Fish & Snodgrass, 2014, 2015, 2019, 2020a, 2020b, 2020c, 2021) regarding their perspectives of online and FTF, the results demonstrated a strong preference by instructors and students to be in the FTF classroom for almost all of the individual and program factors studied. Individual factors studied are specific to the individual (student or instructor), such as motivation, discipline, self-directed learning and independence, schedule flexibility, time and cost investment, preference, happiness and appropriateness for learning environment. Program factors, which are decisions that the instructor makes in developing the course, studied include academic difficulty, academic integrity (cheating), instructor-to-instructor interaction, student-to-instructor interaction, and program technologies (Readers interested in a deeper literature review on the individual and program factors should consult Fish & Snodgrass (2014, 2015, 2016a, 2016b).) With the onset and continuation of the pandemic, most Universities migrated to the online environment. The pandemic forced the University to utilize online courses, and therefore, online course delivery became the standard. There is a need to assess both the student and instructor's perspectives with respect to online education (Shieh, Gummer & Niess, 2008). In this study, after a year of online instruction, we simultaneously surveyed instructors and business students as to their perspectives of online versus FTF education. This paper seeks to explore our preliminary results.

LITERATURE REVIEW

Prior to the pandemic and over the past decades, online education in higher education increased and senior administrators perceived online education to be equivalent to FTF education and concluded that there were no significant differences between online and FTF learning outcomes (Allen & Seamen, 2013). However, questions remain about student perceptions (Baker & Umni, 2018) and instructors perceptions of learning and overall satisfaction between online and FTF instruction. Student performance and perceptions are not always equivalent between the two educational methods. Several research studies have demonstrated that students' preferred FTF education over online education (Fish & Snodgrass, 2014, 2015, 2019, 2020a, 2020b, 2020c, 2021; Tratnik, Urh & Jereb, 2019). In yet another study, at an AACSB accredited business school reported over 65% of students strongly disagreed or disagreed that learning is greater from online courses (Kuzma, Kuzma & Thiewes, 2015). Note these studies took place prior to the pandemic – and most compared the two methods from the FTF perspective. The pandemic has altered every aspect of our lives. Going forward, higher education will undoubtable be different than it was before. The question before all educators and administrators is what education will '*look like*' in the future? While some students clamor for the online environment, others long to be back in the FTF classroom. Therefore, online student perspectives – especially in comparison to traditional FTF education - remain critical to the future of education.

As methods and technology to teach and learn online continues to evolve, research on student perceptions of the online learning environment versus the traditional FTF environment continue (e.g. Allen & Seaman, 2013; Baker & Unni, 2018; Barnes, 2017, Fish & Snodgrass, 2014, 2015, 2019, 2020a, 2020b, 2020c, 2021 In Press; Guest, Rhode, Selvanathan & Soesmanto, 2018, Perreault, Waldman, Alexander & Zhao, 2008; Tanner, Noser, and Langford, 2003; Tanner, Noser, Fuselier & Totaro, 2004a; 2004b; Tanner, Noser, Totaro & Birch, 2006; Tanner et al., 2009; Tratnik, Urh & Jereb, 2019). Our intent with this literature review is not intended to be a comprehensive review of the literature on students' perceptions. Rather, it serves to highlight the ambiguity that exists in the debate between online and FTF education. Some research indicates that online and FTF courses are equally effective (Baker & Unni, 2018; Fowler, 2005; Horspool & Lange, 2012; Topper, 2007). Others show a higher satisfaction for online over FTF environments (Connolly, MacArthur, Stansfield & McLellan, 2007), and yet others, show a preference to FTF over online environments (e.g., Fish & Snodgrass 2014, 2015, 2019, 2020a, 2020b, 2021; Kuzma, Kuzma & Thiewes, 2015; Mullen & Tallent-Runnels 2006; Tratnik et al., 2019). As noted previously (Fish & Snodgrass, 2014, 2015, 2019, 2020a, 2020b, 2020c, 2021), these studies differ in the size (small, medium, large universities), audience (e.g.scientific versus social sciences, business versus non-business, and graduate versus undergraduate), and method of research (e.g. interview, survey). Several online perception studies were completed at large universities or in a public forum (Tanner et al., 2003; Tanner et al., 2004a; 2004b; Tanner et al, 2006; Tanner et al., 2009) or in moderate public forums (Hilton et al., 2020) or in non-business fields (e.g. Dobbs, Waid, & delCarmen, 2009; Lanier, 2006; Leasure, Davis & Theivon, 2000; Reilly et al., 2012; Tekinarslan, 2011; Wang & Morgan, 2008). Studies have explored differences between business and non-business students (Tanner et al., 2004a, 2004b), nursing students (Bilings, Skiba & Connors, 2005), and criminal justice students (Dobbs et al, 2009). Therefore, the context of the study may be an important factor to consider in interpretation of the survey results.

Another researcher noted that studies into instructor and student perceptions will continue to evolve as technology evolves (Richardson, Besser, Koehler, Lim. & Strait, 2016). Information and knowledge regarding instructor beliefs are important to improving instructional effectiveness (Farrell & Kun, 2008). With respect to instructor's perceptions, researchers noted that perceptions vary and have changed over the years. As for effectiveness, a 2004 study at public and nonprofit private institutions revealed that faculty rated online education slightly more effective overall (Guidera, 2004). However, in other studies, instructors perceived online instruction to be inferior to traditional FTF teaching (Fish & Gill, 2009; Wilson, 2001). Similarly, for the individual and program factors studied, the literature review in our prior research noted differences in instructors' perceptions for online or FTF educational methods (Fish & Snodgrass, 2016a, 2016b).

While students at the University have been surveyed twice over this study (Fall 2012 and Spring 2018), instructors have only been surveyed once on these factors in spring of 2017. Survey results demonstrated that over 75% of instructors at the University preferred to teach FTF. Comparison for the individual and program factors showed significant differences in instructor perspectives for instructors who had taught online versus those that had not on every individual and program factor except for cost investment and a slight significance for difficulty (Fish & Snodgrass, 2018 a, 2018b). In general, the instructors at the University of interest favored the FTF environment for most individual and program factors.

Many business student perceptions studies were published over 10 years ago (e.g. Perreault et al., 2008; Tanner et al., 2003; Tanner et al., 2004-1, 2004-2), and similar to other studies (Mortagy & Boghikian-Whitby, 2010; Perreault et al., 2008), perceptions may have changed. In 2012, prior research noted that for all but three individual and program factors, significant differences exist between the online and FTF students' perceptions, and students favored the FTF environment (Fish & Snodgrass, 2014, 2015). The 2018 study reported on significant differences between online and FTF students on all but three factors – interestingly, not the same three factors (Fish & Snodgrass, 2020 a, 2020 b). Specific to the online business students, changes in online student perspectives occurred for self-directed, time investment and appropriateness (Fish & Snodgrass, 2020 a, 2020 b). While most factors did not change in their significance over the six years, for some factors, student perspectives for individual and program factors significant changes were detected (Fish & Snodgrass, 2020 a, 2020b). Essentially both groups had similar perspectives on which environment they favored, as in general, they favored FTF over online but indicated that cheating was easier online and schedule flexibility were preferred online.

While we have been following both the instructor and student perceptions regarding online versus FTF, our literature review did not find any studies that compared these populations perspectives regarding online and FTF education. This study may represent one of the first, and at a minimum, the first to compare instructor and student perspectives during the pandemic. Our study was conducted at a mid-sized, Jesuit, Catholic, business school with a focus on teaching. The research focus lies in uncovering student perceptions where FTF class sizes average 17 students with a capacity of 35 students. At the time of the original surveys, online education was growing, but it was not the primary delivery mechanism at the University as during the pandemic. At the time of the current survey, almost every student at the University had experienced online education and the intent of the larger study was to explore instructor and business student perceptions of online education versus traditional FTF education over time (Fish & Snodgrass, 2014, 2015, 2016a, 2016b, 2019, 2020a, 2020b, 2020c, 2021).

Theoretically, people – whether the instructor or student - should perceive the environments equally and not favor either traditional FTF or online education. The focus of the research presented in this paper seeks to explore the impact of the pandemic on instructor and student perspectives of online and FTF education for individual and program factors. This leads to our Research Question addressed within this paper:

#1) Do instructors and business students' perceptions of online versus Face-to-Face education differ?

- H_{10}: Instructor and business students' perceptions for individual and program factors of online education are *the same* as their perceptions of FTF education.
- H_{11}: Instructor and business students' perceptions for individual and program factors of online education are *not equal* to their perceptions of FTF education.

METHODOLOGY

As part of a larger study at an AACSB-accredited, Jesuit Catholic University in the northeast, instructor and business students received an online Qualtrics-administered survey during April 2021, one year after the start of the COVID pandemic. Each survey (instructor and student) was available over the entire month and in order to increase participation, it was sent to each audience three times over the month through a list serve. The University Internal Review Board and Academic Vice President's granted approval for distribution.

Based upon the above research and insight into the online versus FTF learning environments, the instructors designed a survey on perceptions of online versus FTF learning environments. The instructor survey can be viewed in Appendix A, while the student survey can be accessed at: Student Survey at:http://www.cambriainstitute.com/journals/j.brcacadjb.2015.04.01.wa04.pdf. The wording on each survey was essentially the same with respect to the statements made regarding the individual and program factors; however, slight changes - such as 'taught' versus 'taken', were made to denote the particular audience taking the survey. There were 3 major sections to the survey: demographic questions, Section A for those who had online experience or Section B for those who did not have online experience. The instructor survey asked demographic questions regarding age, gender, respective school the instructor associates with, teaching level (undergraduate, graduate, or both), level taught, faculty rank, self-reported technological skill level, online course experience as a student, and online teaching experience. The student survey asked demographic questions regarding gender, age, class rank (undergraduate – freshmen, sophomore, junior, senior or graduate), undergraduate major or graduate program and potential concentration (graduate), online experience, self-reported level of technological understanding, whether the student was a transfer student, and if the student took an online course, the number of online courses taken. Both surveys then divided into the participants into two groups: individual who have taught (instructors) or taken (students) at least 1 online course completed Section A questions, while individuals who have never taught (instructors) or taken (students) a course in the online environment completed Section B questions. Sections A and B have corresponding questions on the perceptions noted; however, Section A statements are specific to "*I found*" versus Section B statements are "*I perceive*". Specific individual perceptions include individual (student or instructor) perceptions, such as motivation, discipline, self-directed learning and independence, schedule flexibility, time and cost investment, preference, happiness and appropriateness for learning environment, as well as cultural differences. Program factors, which are decisions that the instructor makes in developing the course, studied include: academic difficulty, academic integrity (cheating), instructor-to-instructor interaction, student-to-instructor interaction, and program technologies.

The survey used a five-point Likert scale for each of the factors: significantly dislike, dislike, okay, like, significantly like. The last questions in each section asked the instructor if the instructor or student if they would prefer the opposite environment, the instructor's emotional happiness with the learning environment, and whether the instructor felt that online courses were appropriate for the institution. For instructors or students with online experience, the last open-ended question inquired as to why they chose to offer an online course. For instructors or students without online experience, the survey included an open-ended question inquiring 'why not'.

The liberal arts institution is in a state that requires business students at liberal arts Universities to take 50% of their credit hours in liberal arts. Hence, students would experience courses taught by instructors throughout the University and not just within the School of Business. Also, as a matter of context, it is important to understand that instructors at the University do not use online course designers, are responsible for delivery of online content, and are encouraged to use the University's platform (Desire2Learn) as well as other software and programs (e.g., screen-o-matic, Youtube, Dropbox, Zoom, etc.) in online and FTF course delivery.

ANALYSIS

One hundred thirty instructors and one hundred seventy-nine business students voluntarily participated. Ninety-three instructors and one hundred eleven business students completed section A: online perspective, while 25 instructors and 8 students completed section B: face-to-face perspective. (In spite of the pandemic and switch to 100% online at the end of the spring 2020 semester, the University was able to offer FTF courses in the fall of 2020 and spring of 2021, and therefore, a few students and instructors were able to avoid taking 100% online courses.) Since students and instructors were not required to answer every question, the data set reflected fewer responses for some factors.

Instructors and students' perspectives were compared based upon whether they completed Section A (online) or Section B (FTF). Chi-Square analysis using the contingency coefficient as the nominal value was performed using SPSS. With respect to instructors and business students that had experienced at least one online course, as noted in Table 1, significant differences between instructor and students' perspectives exist for difficulty (p=.000), motivation (p=.007), self-directed (p=.002), independence (p=.003), schedule flexibility (p=.000) and time investment (p=.017). Perspectives on cheating were also slightly significant (p=.085).

As noted in the Appendix B, the majority of students were 'indifferent' on difficulty, but the majority of instructors found online to be 'more' difficult than FTF. While the majority of instructors were 'indifferent' or 'less' motivated online, the majority of students were 'indifferent' to 'significantly less' motivated online. Student responses for self-directed were closely 'equal' for each of the categories, while instructors tended to be 'indifferent' or liked the online environment 'less' or 'significantly less' than FTF. Students enjoyed the independence of the online environment 'more' or 'significantly more' as the majority of instructors tended to enjoy the independence of the online environment 'more' or the 'same' as FTF. Students enjoyed the schedule flexibility 'significantly more', while instructors were slightly less positive as the majority of instructors enjoyed the schedule flexibility 'more' than FTF. While both instructors and students tended to find the online courses take more time than FTF; however, the majority of instructors tend to find the online courses take 'more' time than FTF, the majority of students are more 'indifferent' to the same environment. The majority of instructors perceived that cheating was the ‘same’, ‘easier’ or ‘significantly easier’ to cheat online, while students tended to find it be the ‘same’ or ‘easier’ to cheat online.

Instructors and business students were similar in their perceptions of student-to-student interaction, student-to-instructor interaction, discipline, cost investment, preference for the opposite environment, happiness with the online environment, and appropriateness of online. Instructors and students ‘significantly disliked’, ‘disliked’ or were ‘indifferent’ toward student interaction online. Similarly, the majority of instructors and students ‘disliked’ the student to instructor interaction online. As for discipline, the majority of instructors and students were ‘indifferent’, ‘liked’ or ‘significantly liked’ the discipline required online. The majority of instructors and students are indifferent to the cost differences between online and FTF education. The majority of instructors and students are okay with the online environment and feel that online is appropriate for the institution. Most importantly, *while the respondents are responding from the online perspective, the majority of instructors and students would prefer to take the course in the FTF environment.*

Therefore, the results supported hypotheses H_{10} for some factors and H_{11} for others. For the population of this study, instructors and business students who have experienced online education differ on their perspectives of online education for several factors. However, in general, both groups tend to favor the FTF environment.

Table 1. Chi-Square Analysis based upon Instructor vs Student Online Environment (Section A)

Metric	Pearson Chi-Square Value	Df	Asymptotic Significance (2-sided)	Pearson’s R	Spearman Correlation
Difficulty	23.023	4	.000 *	-.333	-.334
Motivation	14.097	4	.007 *	-.057	-.072
Student-to-Student Interaction	2.931	4	.569	.012	.012
Student-to-Instructor Interact	3.116	4	.539	.022	.008
Discipline	3.204	4	.524	.040	.051
Cheat	8.196	4	.085 **	.161	.156
Self-directed	16.987	4	.002 *	2.14	2.13
Independence	15.879	4	.003 *	.215	.225
Schedule flexibility	27.411	4	.000 *	.233	.270
Time investment	12.007	4	.017 *	-.208	-.193
Cost investment	2.408	4	.661	.090	.089
Preference opposite?	4.097	2	.129	.143	.1425
Happiness with environment	2.923	4	.571	.051	.059
Appropriateness of Online	.533	2	.766	.047	.045

* $p \leq .05$, ** $p \leq .10$

With respect to instructors and business students that had never experienced at least one online course (Section B), as noted in Table 2, significant differences between instructor and students’ perspectives exist for student-to-student interaction, preference to take a class online and happiness with the environment. As noted in Appendix C, most instructors perceived that the student-to-student interaction is ‘significantly more’ in FTF than online, and students tend to be positive, but not as positive, toward student-to-student interaction FTF more than online.

(Interestingly, the online groups felt student-to-student interaction was ‘significantly less’, ‘less’ or were ‘indifferent’ online instead of FTF.) Almost every instructor surveyed FTF did not want to teach online, while students were slightly more indifferent to online education. Instructors were overwhelmingly ‘happy’ or ‘very happy’ in the FTF environment for teaching, while students, who were positive toward FTF, were ‘happy’ to be taught FTF.

Table 2. Chi-Square Analysis based upon Instructor vs Student FTF Environment (Section B)

Metric	Pearson Chi-Square Value	Df	Asymptotic Significance (2-sided)	Pearson’s R	Spearman Correlation
Difficulty	.679	3	.878	-.028	-.023
Motivation	2.298	2	.317	.009	.049
Student-to-Student Interaction	6.448	2	.040 *	-.367	-.315
Student-to-Instructor Interact	2.076	2	.354	-.249	-.249
Discipline	3.441	4	.487	.075	.071
Cheat	.373	2	.830	-.100	-.098
Self-directed	1.661	2	.436	-.189	-.162
Independence	6.230	3	.101	-.349	-.376
Schedule flexibility	4.459	4	.347	-.257	-.247
Time investment	4.716	4	.318	.158	.187
Cost investment	4.723	4	.317	.128	.140
Preference opposite?	6.839	2	.033*	-.453	-.446
Happiness with environment	10.289	2	.006 *	-.529	-.486
Appropriateness of Online	.016	2	.992	.019	.020

* p $\leq$ .05, ** p $\leq$ .10

Instructors and business students that never experienced an online course perceived the environments similarly on most factors. Most instructors and students indicated that FTF was less difficult than online. The majority of instructors felt they perceived that they would be ‘more’ or ‘significantly more’ motivated FTF, while the majority of students felt they perceived that they would be ‘significantly more’ motivated FTF. The majority of instructors perceived that student-to-instructor is ‘significantly more’ in FTF than online, while students perceived a positive impact, but not as positive as instructors, on student-to-instructor interaction. The majority of instructor perceived the discipline required to be ‘the same’, however, students varied in their responses from ‘less’ to ‘more’. The majority of instructors and students perceived that cheating would be ‘harder’ in the FTF environment and that that they would enjoy the self-directed FTF environment more than online. The majority of instructors perceived that they enjoy the independence of the FTF environment ‘more’ or ‘significantly more’ than online, while students were ‘indifferent’ or felt they would enjoy the independence ‘more’ FTF than online. While responses varied, the majority of instructors were ‘indifferent’ to the schedule flexibility, while students tended to be ‘more’ favorable toward the schedule flexibility online than FTF. The majority of instructors perceived FTF classes to take ‘less’ time than online, while students perceived FTF classes to take ‘more’ time than online. Responses on cost investment varied; however, the majority of instructors perceived the two environments as being equal, while students perceived costs would be ‘more’ FTF than online. Most instructors felt that online education was appropriate; however, several instructors who were undecided or responded no to online education. Students’ responses were split across the categories for the question of online appropriateness. So, while slight differences exist in majority preference, these factors were not significantly different between instructors and business students. (Note the small population that remains for this section of the survey may have contributed to the lack of significance.)

Therefore, the results supported hypotheses H_{10} for most factors and H_{11} for a few factors. For the population of this study, instructors and business students who have never experienced online education are very similar in their perspectives of online education. Both groups tend to favor the FTF environment.

DISCUSSION

While technology continues to be incorporated into higher education and most courses moved online due to the continuing pandemic, instructor and business students' perspectives on online versus traditional FTF education at a Jesuit, Catholic AACSB-accredited University still tend to favor FTF education. While research continues on instructor and business student perceptions, this paper represents a comparison between the instructors delivering the education and those receiving it. As noted in the literature review, this may be one of the first to do so. For instructors and business students who have experienced at least one online course, the perceptions for those online are not the same for every factor as difference exist. While a much smaller group, instructors and students who have remained online (the FTF group, Section B) are more homogenous in their perspectives as noted by the very few factors that their perceptions statistically differed on (student-to-student interaction, preference for online and appropriateness of online education). Perhaps these differences may be attributed to the fact that education has been traditionally FTF, and both instructors and students have become accustomed to this educational method.

It is interesting to note that the online instructors and business students' perspectives are statistically different on many factors. Undoubtable, these differences will contribute to differences in perceptions moving forward. Perceptions relate to expectations as well as performance. As online education continues to evolve, these gaps between sender (instructor) and receiver (student) need to be bridged in order to develop a stronger educational system. Individual factors of motivation, self-directed, independence, schedule flexibility and time investment, were significantly different between instructors and students. This result speaks to the different perspective that the individuals in each group have with respect to online versus FTF education. As for program factors, the online instructors and students significantly differed on difficulty and slightly on cheating. Their perspectives on student-to-student interaction and student-to-instructor interaction were not significantly different, which speaks to some agreement between the two groups on the education program method.

It is also important to note the high degree of cohesiveness and agreement between the faculty and the students in the FTF environment, as evidenced by the few differences between those two groups. It may be thought that this cohesiveness forms a foundation for effective teaching and learning and that instructors and students are "happy" in that environment. However, when the switch is made to the online environment – perhaps even mandated – many significant differences between instructors and students appear. And perhaps more importantly, both groups indicate their preferences for FTF. There are many implications that can be drawn from that and this certainly requires further examination. But at the moment, suffice it to say that there is "less agreement" between instructors and students in the online environment as opposed to the FTF. This disconnect might be an artifact of the lightning-fast switch to online as required by the pandemic lockdown. In that situation, there was little time to prepare an effective online format for a class that had been designed to be delivered FTF. There might also have been residual negative feelings as an unintended consequence of this forced change, with the hope that the change was only a "temporary" fix to a once-in-a-lifetime disruption and soon everyone will be back in the classroom where they are "happy". However, there is reason to believe that the growth of online education in higher education will only continue. It is therefore imperative that course designers keep these differences in mind and try to find ways to replicate the cohesiveness of the FTF classroom in the online environment.

As we have proposed in our other research articles (Fish & Snodgrass, 2014, 2015, 2019, 2020a, 2020b, 2020c, 2021), the audience may be a significant difference between the various research studies on this topic. The institution studied is regarded as a 'teaching' institution that has offered FTF instruction for over 150 years. Student and instructor responses in this study highlight the perceptions of each group as a FTF institution – and one that is still developing in the online environment.

CONCLUSIONS

Instructors and business students' perceptions as to online versus FTF education reveal that for this population, traditional FTF education still tends to be favored. While the majority of participants in this study have completed at least one online course at the traditional teaching institution, the instructor and business student perspectives differ on many factors, and in general are more positive toward FTF education. The majority of online instructors and students favored taking their courses FTF instead of online. For those instructors and students who have not experienced an online course, their perspectives are very homogenous as few factors were statistically different between the groups. As the pandemic subsides and technology continues to change, this population's adaptation to online education remains as an issue.

REFERENCES

Allen, I., and Seaman, J. (2013). Changing Course: Ten Years of Tracking Online Education in the United States. *The Sloan Consortium (Sloan-C)*, Retrieved on January 11, 2013 from http://sloanconsortium.org/publications/survey/making_the_grade_2006

Baker, D., and Unni, R. (2018). USA and Asia hospitality and tourism students' perceptions and satisfaction with online learning versus traditional face-to-face instruction. *e-Journal of Business Education & Scholarship of Teaching*, 12(2), 40-54.

Barnes, C. (2017). An Analysis of Student Perceptions of The Quality and Course Satisfaction of Online Courses. *Journal of Higher Education Theory and Practice*, 17(6), 92-98.

Connolly, T.M., MacArthur, E., Stansfield, M. and McLellan, E. (2007). A quasi-experimental study of three online learning courses in computing. *Computers & Education,* 49, 345–59.

Dobbs, R., Waid, C.A., and del Carmen, A. (2009). Students' Perceptions of Online Courses: The Effect of Online Course Experience. *Quarterly Review of Distance Education*, Spring 2009, 10(1), 9-26.

Farrell, T. & Kun, S. (2008). Language policy, language teachers' beliefs, and classroom practices. *Applied Linguistics*, 29(3), 381-403. https://doi.org/10.1093/applin/amm050

Fish, L. A. & Snodgrass, C.R. (2021). Changes in Undergraduate and Graduate Perspectives of Online versus Face-to-Face Education. *The BRC Academy Journal of Business*, Volume 11, Number 1, Print ISN: 215208721 Online ISN: 2152-837X Date April 15, 2021, 147-165. https://dx.doi.org/10.15239/j.brcacadjb.2021.11.01.ja06

Fish, L. A. & Snodgrass, C.R (2020c). A Preliminary Study of Age and Gender and Their Influence on Student Perspectives of Online versus Face-to-Face Education at a Jesuit Institution. *Northeast Decision Sciences Institute Conference Proceedings 2020*, Boston, MA, March 26-29, 2020, pp. 802-830.

Fish, L. A. & Snodgrass, C.R (2020b). Changing Business Student Perceptions of Program Factors in Online versus Face-to-Face Education. *Business Education Innovation Journal*, June 2020, pp. 123-131.

Fish, L. A. & Snodgrass, C.R. (2020a). A Preliminary Study of Changing Business Student Perceptions of Individual Factors in Online Versus Face-to-Face Education." *The BRC Academy Journal of Education* 8, no. 1 (2020): 61-83.
https://dx.doi.org/10.15239/j.brcacadje.2020.08.01.ja03
Web Appendix: A web appendix for this paper is available at:
https://dx.doi.org/10.15239/j.brcacadje.2020.08.01.wa03

Fish, L. A. & Snodgrass, C.R. (2019). A Preliminary Study of Changing Business Student Perceptions of Program Factors in Online versus Face-to-Face Education. *Northeast Decision Sciences Institute 2019*, Philadelphia, PA, April 4-6, 2019.
https://nedsi2019.exordo.com/programme/session/47

Fish, L. A. & Snodgrass, C.R. (2018a). "Instructor Perspectives of Online versus Face-to-Face Education at a Jesuit Institution: Program Factors." *The BRC Academy Journal of Education* 7, no. 1: 55–77. http://dx.doi.org/10.15239/j.brcacadje.2018.07.01.ja03
WEB APPENDIX A web appendix for this paper is available at:
http://dx.doi.org/10.15239/j.brcacadje.2018.07.01.wa03

Fish, L. A. & Snodgrass, C.R (2018b). "Instructor Perspectives of Online versus Face-to-Face Education at a Jesuit Institution: Instructor Factors." *The BRC Academy Journal of Education* 7, no. 1: 37–53. http://dx.doi.org/10.15239/j.brcacadje.2018.07.01.ja02
WEB APPENDIX A web appendix for this paper is available at:
http://dx.doi.org/10.15239/j.brcacadje.2018.07.01.wa02

Fish, L. A. & Snodgrass, C.R. (2016a). A Literature Review and Directions for Future Research on International Student Perceptions of Online versus Face-to-Face Education: Student-centered Characteristics. *The BRC Academy Journal of Education*, 5(1), 1-22. http://dx.doi.org/10.15239/j.brcacadje.2016.0501.ja01

Fish, L. A. & Snodgrass, C.R. (2016b). A Literature Review and Directions for Future Research on International Student Perceptions of Online versus Face-to-Face Education: Program-centered Characteristics. *The BRC Academy Journal of Education*, 5(1), 23-52. http://dx.doi.org/10.15239/j.brcacadje.2016.0501.ja02

Fish, L. A. & Snodgrass, C.R. (2015). Student Perceptions of Online versus Face-to-Face Education: Student Characteristics. *Business Education Innovation Journal*, 7(2), 83-94.

Fish, L. A. & Snodgrass, C.R. (2014). A Preliminary Study of Business Student Perceptions of Online versus Face-to-Face Education. *BRC Journal of Advances in Education,* pp. 1-21. DOI: http://dx.doi.org/10.15239/j.brcacadje.2014.04.01.ja01

Fowler, D. (2005). Are on-site courses as effective as online? *Online Cl@ssroom: Ideas for Effective Online Instruction*, March, 1–2.

Guidera, S, G, (2004), Perceptions of the effective-ness of online instruction in terms of the seven principles of effective undergraduate education. *Journal of Educational Technology Systems*, Í2(2& 3), 139-178.

Guest, R., Rhode, N., Selvanathan, S., & Soesmanto, T. (2018). Student satisfaction and online teaching. *Assessment & Evaluation in Higher Education*, 43(7), 1084-1093.

Hilton, R. Moos, C. and Barnes, C. (2020). A Comparative Analysis fo Students' Perceptions of Learning in Online Versus Traditional Courses. *e-Journal of Business Education & Scholarship of Teaching*, 14(3), December 2020, pp. 2-11.

Horspool, A. & Lange, C. (2012). Applying the scholarship of teaching and learning: student perceptions, behaviors and success online and face-to-face, *Assessment & Evaluation in Higher Education*, February 2012, 37(1), 73-88, Accessed on January 8, 2013 from http://dx.doi.org/10.1080/02602938.2010.496532.

Kuzma, A., Kuzma, J., and Thiewes, H. (2015). Business Student Attitudes, Experience, and Satisfaction with Online Courses. *American Journal of Business Education*, 8(2), 121-310.

Lanier, M. (2006). Academic Integrity and Distance Learning. *Journal of Criminal Justice Education*, Sep 2006, 17(2), 244-21.

Leasure, A. R., Davis, L., & Thievon, S. L. (2000). Comparison of student outcomes and preferences in a traditional vs. World Wide Web-based baccalaureate nursing research course. *Journal of Nursing Education, 39*(4), 149-154.

Mortagy, Y. & Boghikian-Whitby, S. (2010). A Longitudinal Comparative Study of Student Perceptions in Online Education. *Interdisciplinary Journal of E-Learning and Learning Objects*, 6, 23-46.

Mullen, G.E., & Tallent-Runnels, M.K. (2006). Student outcomes and perceptions of instructors' demands and support in online and traditional classrooms. *Internet and Higher Education,* 9, 257–66.

Perreault, H., Waldman, L., Alexander, M. & Zhao, J. (2008). Graduate Business Students' Perceptions of Online Learning: A Five Year Comparison. *The Delta Pi Epsilon Journal,* Fall 2008, L(3), 164-179.

Reilly, J.R., Gallager-Lepak, S. & Killion, C. (2012). Me and My Computer: Emotional Factors in Online Learning, *Nursing Education Perspectives*, March/April, 33(2), 100 - 105.

Richardson, J.C., Besser, E. Koehler, A., Lim, J. & Strait, M. (2016). Instructors' Perceptions of Instructor Presence in Online Learning Environments. *International Review of Research in Open and Distributed Learning*, 17(4), 82-104.

Shieh, R.S., Gummer, E. & Niess, M. (2008). The Quality of a Web-Based Course: Perspectives of the Instructor and the Students. *TechTrend*s, November/December, 52(6), 61-68. https://doi.org/10.1007/s11528-008-0220-3

Tanner, J.R., Noser, T.C., & Totaro, M.W. (2009). Business Faculty and Undergraduate Students' Perceptions of Online Learning: A Comparative Study. *Journal of Information Systems Education*, Spring 2009, 20(1), 29-40.

Tanner, J., Noser, T., Totaro, M., & Birch, R. (2006). Student Perceptions of The Online 'Classroom': An Update. *International Business & Economics Research Journal*, 5(10), 31-38.

Tanner, J., Noser, T., Fuselier, J., & Totaro, M. (2004a). 'The Online 'Classroom': Differences in Perception between Business Students and Non-Business Students. *Journal of College Teaching and Learning*, 1(3), 37-44.

Tanner, J., Noser, T., Fuselier, J., & Totaro, M. (2004b). 'The Online 'Classroom': What Do Students Think*? Journal of Informatics Education Research*, 6 (1), 43-54.

Tanner, J., Noser, T., & Langford, H. (2003). Perceptions of Undergraduate Business Students Toward Online Courses In Higher Education Expanded and Revisited: Do Gender, Age, and/or Past Experiences Make a Difference? *Journal of Business and Economics Research*, 1(2), 13-20.

Tekinarslan, E. (2011). Faculty of Education Students' Self-efficacy Perceptions toward Online Technologies. *Electronic Journal of Social Sciences*, Summer 2011, 10(37), 120-134.

Topper, A. (2007). Are they the same? Comparing the instructional quality of online and face-to-face graduate education courses. *Assessment & Evaluation in Higher Education,* 32 (6), 681–691.

Tratnik, A., Urh, M. & Jereb, E. (2019). Student satisfaction with an online and a face-to-face Business English course in a higher education context. *Innovations in Education and Teaching International*, 56(1), 36-45.

Wang, L.C. & Morgan, W.R. (2008). Student Perceptions of Using Instant Messaging Software to Facilitate Synchronous Online Class Interaction in a Graduate Teacher Education Course. *Journal of Computing in Teacher Education*, Fall 2008, 25(1), 15-21.

Wilson, C. (2001). Faculty attitudes about distance learning. *Educause Quarterly*, 2, 70-71.

APPENDIX A
Instructor Perceptions Survey
As part of a continuing study into perspectives surrounding online education, we are requesting your input into a survey that compares online and face-to-face instructional perspectives. Please answer the following questions:

Background Information:

What school do you teach in?	School of Arts and Sciences (1) School of Education and Human Services (2) School of Business (3)
How old are you?	
Are you:	Male (1) Female (2)
What is your rank at the University?	Lecturer/ Adjunct Professor (1) Assistant Professor (2) Associate Professor (3) Professor (4) Other (5)
What is your highest education level?	Associates (1) Bachelors (2) Masters (3) Doctorate (4) Post-Doctorate (5)
Do you teach undergraduate courses only, graduate courses only, or both?	Undergraduate courses only (1) Graduate courses only (2) Both undergraduate and graduate courses (3)
Would you classify your technology understanding as:	• Do not use technology in the classroom as part of teaching or to communicate with students. (1) • Beginner/Novice (e.g. use Microsoft Office Powerpoint (or similar projection software) or Excel as part of teaching and email, but no other software to conduct class or communicate with students). (2) • Advanced Beginner (e.g. use 'some' Powerpoint (or similar projection software), some Desire2Learn features, Microsoft Office or similar to conduct class and communicate with students). (3) • Intermediate (e.g. Use 'many' Desire2Learn features in class, Microsoft Office, and other computer applications (R, SPSS, etc.) to conduct class and communicate with students).). (4) • Extensive (e.g extensive knowledge and use of various computer software programs to conduct class and communicate with students). (5)
How many years have you taught at the University?	
How many years have you taught at the University level (including the University)?	

Have you ever taken an online course at the University (including the online instructor training course)?	Yes (1) No (2)
Have you ever taken an online course at a school other than at the University?	Yes (1) No (2)
Have you taught an online course?	Yes (SECTION A) (1) No (SECTION B) (2)

SECTION A - AT LEAST ONE ONLINE COURSE AT THE UNIVERSITY

How many online courses have you taught?	
Have you taught an online course at a school other than at the University?	Yes (1) No (2)
How many online courses have you taught at a school other than this University?	
Prior to teaching an online course, did you take a course to prepare you for the online environment?	Yes (1) No (2)
If 'yes', was the course offered by the University or the book publisher?	University(1) Book Publisher(2) Other (3)
With respect to teaching online courses at Canisius College compared to face-to-face (traditional classrooms), please rate the following responses:	
I find online courses to be______in difficulty than face-to-face courses.	(1) Significantly Easier (2) Easier (3) The Same Difficulty (4) Harder (5) Significantly Harder
I am _____ motivated in online courses than face-to-face courses.	(1) Significantly Less (2) Less (3) Equally (4) More (5) Significantly More
I _____the interaction between students in the online environment compared to the face-to-face course environment.	(1) Significantly Dislike (2) Dislike (3) Equate (4) Like (5) Significantly Like
I _________ the interaction between the instructor and students in the online environment compared to the face-to-face environment.	(1) Significantly Dislike (2) Dislike (3) Equate (4) Like (5) Significantly Like
I find the discipline required in taking online courses to be ____ than in face-to-face courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More

I find that it is ____ for students to cheat in an online course than in face-to-face courses. .	(1) Significantly Easier (2) Easier (3) The Same (4) Harder (5) Significantly Harder
I enjoy the self-directed online learning environment__than the interaction in face-to-face courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I enjoy the independence associated with the online learning environment_____than the interaction in face-to-face courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I enjoy the schedule flexibility associated with the online learning environment_____than the interaction in face-to-face courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I find online courses require_____time investment in the course than face-to-face courses	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I find online courses total costs are _____than face-to-face courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
In the online environment, I feel these activities increase my students' understanding of the course material. (Check all that apply.)	(1) Additional Readings (Not including textbook) (2) Assignments/Homework (3) Course Surveys (4) Discussion boards (5) In-class sessions (live sessions which may include team or group work) (6) Instructor live lectures (7) Instructor taped lectures (8) Instructor Office Hours (9) Instructor Posted Notes (10) Laboratory/ experiential activities (11) Other students (12) Problem Hints & Scaffolding Examples (13) Textbook (14) Course related videos

	(15) Other__________
In the online environment, I feel these activities decrease my students' understanding of the course material. (Check all that apply.)	(1) Additional Readings (Not including textbook) (2) Assignments/Homework (3) Course Surveys (4) Discussion boards (5) In-class sessions (live sessions which may include team or group work) (6) Instructor live lectures (7) Instructor taped lectures (8) Instructor Office Hours (9) Instructor Posted Notes (10) Laboratory/ experiential activities (11) Other students (12) Problem Hints & Scaffolding Examples (13) Textbook (14) Course related videos (15) Other__________
Would you prefer to teach the class in a traditional face-to-face environment?	Yes (1) Undecided(2) No (3)
I am ____with the online course environment for teaching.	(1) Not very happy (2) Not happy (3) Okay (4) Happy (5) Very happy
Given this institution, do you think online courses are appropriate?	Yes (1) Undecided(2) No (3)
Why did you choose to teach the course online?	

Thank you for your time! It is greatly appreciated.

SECTION B - NEVER TAKEN AN ONLINE COURSE AT THE UNIVERSITY.

With respect to teaching face-to-face courses (traditional) at the University compared to teaching online, please rate the following responses: :	
I perceive face-to-face courses to be______in difficulty than online courses.	(1) Significantly Easier (2) Easier (3) The Same Difficulty (4) Harder (5) Significantly Harder
I perceive that I would be _____ motivated in face-to-face courses than online courses.	(1) Significantly Less (2) Less (3) Equally (4) More (5) Significantly More

I perceive that I would _______ the interaction between students in the face-to-face environment compared to the online environment.	(1) Significantly Dislike (2) Dislike (3) Equate (4) Like (5) Significantly Like
I perceive that I would __________ the interaction between the instructor and students in the face-to-face environment compared to the online environment.	(1) Significantly Worse (2) Worse (3) Equal (4) Better (5) Significantly Better
I perceive the discipline required in taking face-to-face courses to be_____than in online courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I perceive that it would be _____ to cheat in face-to-face courses than in online courses.	(1) Significantly Easier (2) Easier (3) The Same (4) Harder (5) Significantly Harder
I believe that I would enjoy the self-directed face-to-face environment ______ than the interaction in online environment.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I believe that I would enjoy the independence associated with face-to-face courses _____ than the interaction in online courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I believe that I would enjoy the schedule flexibility associated with face-to-face courses ____ than in online courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I believe that face-to-face courses require _____ time investment in the course than online courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
I believe face-to-face courses total costs are _____than online courses.	(1) Significantly Less (2) Less (3) The Same (4) More (5) Significantly More
In the face-to-face environment, I feel these activities increase my students' understanding of the course material. (Check all that apply.)	(1) Additional Readings (Not including textbook) (2) Assignments/Homework

	(3) Course Surveys (4) Discussion boards (5) In-class sessions (live sessions which may include team or group work) (6) Instructor live lectures (7) Instructor taped lectures (8) Instructor Office Hours (9) Instructor Posted Notes (10) Laboratory/ experiential activities (11) Other students (12) Problem Hints & Scaffolding Examples (13) Textbook (14) Course related videos (15) Other__________
In the face-to-face environment, I feel these activities decrease my students' understanding of the course material. (Check all that apply.)	(1) Additional Readings (Not including textbook) (2) Assignments/Homework (3) Course Surveys (4) Discussion boards (5) In-class sessions (live sessions which may include team or group work) (6) Instructor live lectures (7) Instructor taped lectures (8) Instructor Office Hours (9) Instructor Posted Notes (10) Laboratory/ experiential activities (11) Other students (12) Problem Hints & Scaffolding Examples (13) Textbook (14) Course related videos (15) Other__________
Would you prefer to teach the class in an online environment?	Yes (1) Undecided(2) No (3)
If not, why not?	
I am ____with the face-to-face environment for teaching.	(1) Not very happy (2) Not happy (3) Okay (4) Happy (5) Very happy
Given this institution, do you think online courses are appropriate?	Yes (1) Undecided(2) No (3)

Thank you for your time! It is greatly appreciated.

APPENDIX B
Online (Section A) Instructor and Student Number of Responses on Significant and Slightly Significant Factors

Difficulty	I find online courses to be ____________ in difficulty than face-to-face courses.					Total
	1 Significantly easier	2 Easier	3 The same	4 Harder	5 Significantly Harder	
Instructor	2	6	29	**37**	18	92
Student	10	24	**42**	28	7	111
Total	12	30	71	65	25	203

Motivation	I am ________ motivated in online courses than face-to-face courses.					Total
	1 Significantly less	2 Less	3 Equally	4 More	5 Significantly More	
Instructor	12	21	**50**	6	4	93
Student	**30**	**23**	**35**	14	9	111
Total	42	44	85	20	13	204

Cheat	I find that it is ____ for students to cheat in an online course than in face-to-face courses.					Total
	1 Significantly easier	2 Easier	3 The same	4 Harder	5 Significantly Harder	
Instructor	**26**	**29**	**34**	1	1	91
Student	16	**40**	**48**	6	1	111
Total	42	69	82	7	2	202

Self-Directed	I enjoy the self-directed online environment ______ than the interaction in face-to-face environment.					Total
	1 Significantly Less	2 Less	3 The same	4 More	5 Significantly More	
Instructor	22	24	**30**	9	7	92
Student	23	15	25	25	23	111
Total	45	39	55	34	30	203

Independence	I enjoy the independence associated with the online courses ______ than the interaction in face-to-face courses.					Total
	1 Significantly Less	2 Less	3 The same	4 More	5 Significantly More	
Instructor	14	19	24	**30**	6	93
Student	12	13	24	32	**30**	111
Total	26	32	48	62	36	204

Schedule Flexibility	I enjoy the schedule flexibility associated with online courses______ than the interaction in face-to-face courses.					Total
	1 Significantly Less	2 Less	3 The same	4 More	5 Significantly More	
Instructor	6	18	14	**39**	15	92
Student	9	3	16	35	**48**	111
Total	15	21	30	74	63	203

Time Investment	I find online courses require ______ time investment in the course than face-to-face courses.					Total
	1 Significantly Less	2 Less	3 The same	4 More	5 Significantly More	
Instructor	1	2	26	**39**	24	92
Student	6	13	**36**	34	22	111
Total	7	15	62	73	46	203

APPENDIX C
FTF (Section B) Instructor and Student Number of Responses on Significant and Slightly Significant Factors

Student-to-Student Interaction	I perceive that I would _______ the interaction between students in the face-to-face environment compared to the online environment.					Total
	1 Significantly Dislike	2 Dislike	3 Equate	4 Like	5 Significantly Like	
Instructor			1	8	**16**	25
Student			3	2	3	8
Total			4	10	19	33

Preference Opposite?	Would you prefer to teach (take) the class in an online environment?			Total
	1 Yes	2 Undecided	3 No	
Instructor	0	1	24	25
Student	1	2	5	8
Total	1	3	29	33

Happiness	I am _____ with the face-to-face environment for teaching (learning).					Total
	1 Not Very Happy	2 Not Happy	3 Okay	4 Happy	5 Very Happy	
Instructor			0	9	16	25
Student			2	4	1	7
Total			2	13	17	32

DECISION SCIENCES INSTITUTE

Pandemic-Driven Educational Adaptation and Four Practices to Carry into the Post-Pandemic Classroom

Alanah Mitchell
Department of Information Management and Business Analytics
College of Business and Public Administration
Drake University
Email: alanah.mitchell@drake.edu

ABSTRACT

Teaching and learning during the COVID-19 pandemic presented both challenges and opportunities for faculty to rethink their course plans and implementation. This work looks to learn from teaching during the pandemic with the presentation of four key practices that were successful and can be continued post-pandemic. An informal survey of valuable pandemic practices, combined with research support, provides a foundation for discussion and speculation on how faculty might move forward post-pandemic.

KEYWORDS: Information Technology Education, Higher Education, COVID-19, Post-Pandemic

INTRODUCTION

When COVID-19 was declared a global pandemic by the World Health Organization (WHO) in March 2020, higher educational institutions were rapidly forced to shift their normal practices in reaction to this crisis (Sharadgah & Sa'di, 2020). This transition not only presented a challenge to universities, faculty, and students, but it also presented an opportunity to rethink traditional educational methods in order to adapt to the online, remote, or hybrid environment forced by the pandemic (Xie et al., 2020).

While the transition in March 2020 required a rapid, emergency response, the semester break over the summer provided some faculty time for planning and preparation, allowing for a more thoughtful course transition for the 2020-2021 academic year. However, teaching and learning during the unprecedented and unpredictable pandemic could be equated to building an airplane while flying as universities were constantly changing policies and requirements related to the pandemic over the course of the year (e.g., Sharadgah & Sa'di, 2020).

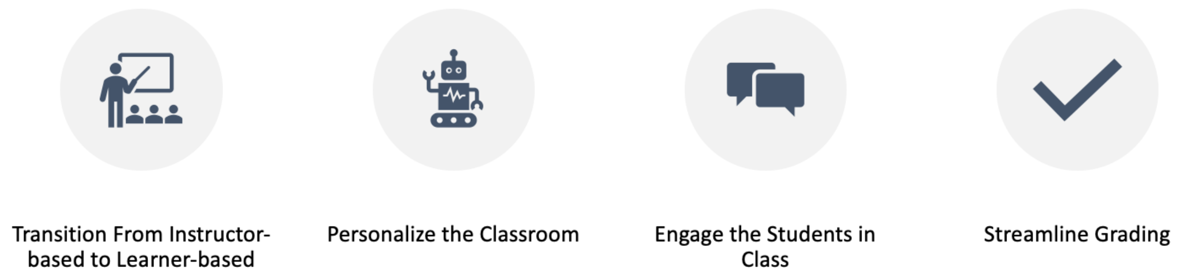
While many students and faculty are looking forward to the post-pandemic future, it is useful to reflect on the lessons learned over the course of the 2020-2021 academic year. Hundreds of hours were spent recording videos, designing engaging online activities, rethinking assignments and assessment, among other activities. Certainly, the teaching adaptation that took place due to the pandemic was not a total loss and there are lessons to be learned from this experience that can be taken forward and applied in future semesters. While the field of education is often considered to be resistant to change (e.g., Blin & Munro, 2008), it would be unfortunate to abandon the lessons learned from the pandemic and relapse to old patterns (Abcouwer et al., 2021). For faculty who strive to innovate and improve their courses, implementing lessons learned from the pandemic, post-pandemic, is a logical step.

Based on this background, the goal of this work is to identify valuable practices from teaching and learning during the pandemic that can be applied in post-pandemic semesters. Specifically, this paper poses the following research question asking: *What valuable aspects from teaching during the pandemic should be adopted in future semesters?* To answer this question, an informal survey was conducted with 15 faculty that taught during COVID-19, representing the disciplines of business analytics, information systems, operations management, and quantitative analysis. The following sections present the four primary practices that emerged from this informal survey integrated with existing related research.

PRACTICES TO CARRY INTO THE POST-PANDEMIC CLASSROOM

In order to address the research question from this work and to identify aspects from teaching during the pandemic that should be adopted post-pandemic, faculty were asked a handful of questions about their pandemic teaching plans and experiences. Specifically, survey questions asked: 1) what did you love about teaching in the time of the pandemic? and 2) what (if any) valuable practices from teaching during the pandemic do you want to continue in the future? Faculty feedback from this survey ranged from classroom management techniques (e.g., engagement, short videos, etc.) to specific assignments and activities (e.g., paired programming, Jeopardy, etc.). Ultimately, the feedback revealed key ideas or themes in four main areas including: 1) transition from instructor-based to learner-based, 2) personalize the classroom, 3) engage the students in class, and 4) streamline grading. Figure 1 highlights the four practices from this review worth continuing post-pandemic, while the following sections discuss each idea further including support from prior research.

Figure 1: Practices to Carry into the Post-Pandemic Classroom



Practice #1: Transition from Instructor-Based to Learner-Based

There are numerous well-known theories of learning that have been studied over time including the objectivist model, the constructivist model, the cooperative (or collaborative) model, the cognitive information processing model, and even the sociocultural model (Leidner & Jarvenpaa, 1995). While the objectivist model of learning is a more traditional model emphasizing the idea that teachers should pass on knowledge to students, some of the other models (e.g., constructivist and cooperative) are more learner focused, rather than instructor focused, suggesting that students benefit from information discovery on their own or in a group (Leidner & Jarvenpaa, 1995). The pandemic classroom certainly challenged the traditional objectivist model of learning as students could not be in a classroom with their professors like they were used to (or if they were co-located, they were six feet apart).

Therefore, in order to adapt during the pandemic, some surveyed faculty experimented with their classroom learning models and found success with a flipped format course structure, which is a more learner-based approach. In fact, this structure was adopted and summarized in the literature as a useful pandemic strategy (e.g., Connolly & Mutchler, 2021; Toney et al., 2021). Some adaptations of the flipped format included short videos online, allowing class time for questions, individual work, group work, coding, discussions, and quizzes (Toney et al., 2021). Others used online course periods for live lectures, with in person class time used for coding demonstrations, class activities, assignments, and question-and-answer time (Connolly & Mutchler, 2021). In each of these cases, the goal of this type of course design asks students to do activities outside of class time (e.g., read and watch videos). Supplementary materials and information available online (e.g., electronic databases, e-books, data) can even be made accessible to students to enhance their self-paced learning (Xie et al., 2020). As an added bonus, some of the out of class activities can be designed for long term usage. For instance, Sanandaji and Ghanbartehrani (2021) designed a flipped technology course with the intention of usage both during and post-pandemic. A few of the faculty surveyed found that their students really liked and appreciated having online videos available outside of class time, and they plan to continue making videos available post-pandemic. Notably, one faculty member discussed the importance of ensuring online materials emphasize accessibility (e.g., live transcripts, captioning, ease of code copying, etc.). Ultimately, with a flipped format requiring some work done outside of class, in person class time can be freed up for other activities (e.g., problem solving and application).

In order for students to be successful in a learner-based course structure, students need to be self-motivated, which can be a challenge (Xie et al., 2020). However, the use of a learning management system (LMS) as the “single source of truth” can be used to organize course materials into modules helpful for clarity and communication (Connolly & Mutchler, 2021, p. 4). Additionally, the use of a pacing guide can be provided to students with a plan of activities for each day in order to assist with the development of time management skills (Hvalshagen et al., 2021). Frost (2021) even addressed the idea of course structure and scaffolding in his COVID-19 classroom by combining theory and application in videos as well as through the use of daily email updates and reminders including due dates, hints, theory nuggets, and request for help. Indeed, one faculty member surveyed found that the use of regular email communication really helped to keep students on track. Ultimately, when using a learner-based model, like the flipped classroom model, clear communication of the class plans and schedule is critical (Connolly & Mutchler, 2021; Toney et al., 2021).

Going forward, teaching and learning can continue to be both synchronous and asynchronous (indeed, it should have always been this way). Faculty should consider how to take advantage of the best of face-to-face and the best of online in order to design hybrid courses that achieve the benefits of both approaches. In the post-pandemic world, some organizational leaders are considering hybrid work for their employees (McGregor, 2021), suggesting student preparation to work in a hybrid environment may be an added benefit with this approach.

Practice #2: Personalize the Classroom

Research recognizes the role of social connection and a physical campus as beneficial for collaboration and relationship development and unfortunately, during the pandemic, some learners felt very distant from their campus, instructors, and peers (Xie et al., 2020). Indeed, this challenge was identified in the faculty survey from this study. However, the opposite viewpoint was also reflected. In fact, a few faculty commented that they felt closer to students as they had seen where they lived, shared pet sightings and stories, and generally felt more personal connections than pre-pandemic. Notably, some faculty reported that students demonstrated more consideration and understanding during the pandemic (Chen & Roldan, 2021).

To build on the consideration and understanding from the pandemic, faculty surveyed identified a few practices they would like to continue post-pandemic including using student names (made easier in thanks to Zoom), having accessible virtual office hours (again, thanks to Zoom), and using technology to invite questions and check the temperature of the room (e.g., using Slido, PollEverywhere, Mentimeter, or the like). Research from Dick (2021) outlined even more ideas for establishing rapport and communicating with students as a means for increasing classroom engagement including introductions with the goal of humanizing the professor and students, using student names, prompt responses, encouragement of office visits, and other ideas focused on making students feel special. Some faculty even found success by providing their phone numbers for students to text (e.g., Dick, 2021; Frost, 2021). It is worth noting that Remind.com is an option for texting that does not require revealing one's phone number.

Post-pandemic, it would be beneficial to continue the formation and development of personalized relationships among faculty and students. Kaufman and Schipper (2018) outline a number of ideas for developing the skills necessary to teach with the emotional, social, and intellectual well-being of students in mind and may be a useful resource for continued development in this area. As was the case during the pandemic, faculty need to continue to be flexible, reasonable, and transparent with students (Connolly & Mutchler, 2021). Faculty should work to let go of the idea of perfectionism, continue to have empathy for themselves and their students, and enjoy the evolution of their courses.

Practice #3: Engage the Students in Class

All of the faculty surveyed were looking forward to post-pandemic, face-to-face, mask-less interactions in their classrooms. From a research perspective, humans need to connect to one another (Aleman & Sommer, 2020) and as an impact of the pandemic, even in cases where students were meeting synchronously, there were still challenges with actually "seeing" one another and connecting (Castelli & Sarvary, 2021).

In an effort to design engaging class time, the surveyed faculty tried out a number of different techniques. Some of these ideas including playing games (e.g., Jeopardy or Kahoot) as well as working in groups (e.g., code sharing, paired programming, breakout groups, and team assignments). There is much research that supports the use of collaborative work for learning, suggesting students are all starting from different points but faculty can identify goals and objectives that recognize student diversity and allow them to cooperatively work together (Abcouwer et al., 2021). Active learning is a well-known methodology that involves the engagement of students both doing and thinking about things in the learning environment (Bonwell & Eison, 1991) and has been found to be a useful predictor of student success (Serva & Fuller, 1999). Prior research has identified dozens of options for active learning assignments within the discipline of information technology including the use of demos, cases, simulations, and inviting guest speakers, to name a few (Mitchell et al., 2017). With the use of video conferencing tools, guest speakers and recent alumni can be invited from other geographical locations as a way to personalize interactions even more (e.g., Olsen, 2021).

While increasing the engagement in the classroom can make attending class more entertaining, it can also help with the professional preparedness of students. For example, some research has suggested technical skills alone are insufficient for students in the field of information technology. Students need to work to develop soft skills like teamwork and communication (Osmani et al., 2016), recognizing an increased interest in the development of social competence from the employer perspective so new hires can work in professional teams effectively post-graduation (Beard et al., 2008; Figl, 2010; Osmani et al., 2016). Prior work recognizes the value of increased student engagement and teamwork in the classroom as a way to prepare students for their future workplaces (Gomez et al., 2009).

Post-pandemic, faculty are going to have to work to emphasize the value of actually being in a classroom, as students now know that it’s possible to produce work without being in a physical classroom. In order to sell the value of class time, faculty should ensure that class time is meaningful. This can likely be accomplished in some of the ways that were tested during the pandemic, including individual work, group work, team brainstorming and problem solving, discussions, question and answer sessions, guest speakers, and other active learning techniques. Ultimately, faculty may need to prepare course schedules that require student engagement in every face-to-face class meeting.

Practice #4: Streamline Grading

Student assessment has historically been considered an important part of the educational experience (Lowell, 1926). Assessment is used by faculty to identify evidence of student learning and can be either formative (e.g., continuous feedback through discussion, verbal response, quizzes, activities, or assignments) or summative (e.g., long term feedback at the end of a learning process) (Sharadgah & Sa’di, 2020). From the student perspective, assessment is often perceived as unpleasant and disruptive (Wu et al., 2008). From the faculty perspective, assessment is often considered time consuming and unenjoyable (Nilson, 2015). The COVID-19 pandemic further complicated assessment goals as several the faculty surveyed were concerned about student cheating due to remote learning.

In an attempt to address assessment and grading concerns during the pandemic, some faculty redesigned their courses to adopt automated multiple choice assessment methods to illustrate concept familiarity through low stakes quizzes and grades. Additionally, faculty relied on the use of objective or subjective questions, scenarios, or problem-solving assignments and exams for high stakes grades within a course. In fact, one study suggested rethinking assignments and exams so that students can show concept understanding as opposed to simply getting the right answer as a means to curb possible cheating (Connolly & Mutchler, 2021). While the use of objective and subjective assessment methods would likely require more grading efforts from faculty, faculty surveyed successfully adopted the use of grading rubrics within their university's LMS to streamline grading. Indeed, using a LMS to post clear grading rubrics allows for students to see what is expected of them in terms of how (but not *what*) to answer and can make grading easier for instructors (Connolly & Mutchler, 2021). Relatedly, specifications grading (or specs grading) extends the use of rubrics in a two-level approach emphasizing student mastery of a topic (e.g., satisfactory or not) thus streamlining grading even more (Nilson, 2015).

Moving forward, faculty surveyed were interested in continuing to adopt assessment and grading practices that cover all levels of Bloom's taxonomy (e.g., remembering, understanding, applying, etc.) (Bloom et al., 1956) by including automated low stakes grading as well as objective and subjective high stakes grading using rubrics to simplifying the grading process.

CONCLUDING COMMENTS

The goal of this paper was to reflect on the 2020-2021 academic year and identify any valuable aspects from teaching during the time of COVID-19 that should be adopted in future post-pandemic, semesters. An informal survey of faculty was used to identify four key practices worth carrying into the post-pandemic classroom. The identified practices broadly include: 1) transition from instructor-based to learner-based, 2) personalize the classroom, 3) engage the students in class, and 4) streamline grading. Previous research was included to support the recommended ideas from this work. Table 1 summarizes each of the four practices from this review and includes possible ideas for implementation related to each practice.

Table 1: Practices to Carry into the Post-Pandemic Classroom and Ideas for Implementation

	Post-Pandemic Practice	Ideas for Implementation
1	Transition from instructor-based to learner-based	• Organize, use, and update course LMS with a clear schedule and course structure (e.g., modules). • Carefully consider which course activities need to be done synchronously and what can be done asynchronously. • Use technology to accommodate accessibility needs (e.g., captioning, lecture transcripts, screen reader support, multiple communication channels, etc.). • Design course materials for re-use.
2	Personalize the classroom	• Humanize oneself; let the students get to know you. • Get to know students, care about their learning as well as their well-being. • Make office hours accessible and encourage students to use them. • Repeatedly check in and ask for and act on the feedback.
3	Engage the students in class	• Sell the value of being in the classroom by making class time meaningful. • Include in-class discussions, activities, exercises, and assignments in every class. • Use groupwork to develop the soft skills of students as well as to create social opportunities for them. • Use technology to compliment in-class activities and develop student skills.
4	Streamline grading	• Use both formative and summative student assessment methods. • Use automated assessments for low stakes work outside of class time. • Design high stakes assignments to emphasize concept application. • Develop assignment rubrics to make grading easier and faster.

What this paper suggests is that the faculty teaching and learning efforts and adaptation that took place during the COVID-19 pandemic were not all for naught. Furthermore, the exploration of valuable pandemic classroom practices from this work contributes to both academia and research going forward. Specifically, in terms of implications, faculty and researchers can use the practices and recommendations identified in this work in the development of plans for future coursework as well as for the identification of possible areas of future research and exploration.

REFERENCES

Abcouwer, T., Takács, E., & Solymosy, J. (2021). Fine-tuning the evaluation focus in the university cooperative learning model in relation to the pandemic. *Communications of the Association for Information Systems*, *48*, 196–204. https://doi.org/10.17705/1CAIS.04825

Aleman, A., & Sommer, I. (2020). The silent danger of social distancing. *Psychological Medicine*, 1–2. https://doi.org/10.1017/S0033291720002597

Beard, D., Schwieger, D., & Surendran, K. (2008). Integrating soft skills assessment through university, college, and programmatic efforts at an AACSB accredited institution. *Journal of Information Systems Education*, *19*(2), 229–240.

Blin, F., & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance to change through the lens of activity theory. *Computers & Education*, *50*, 475–490. https://doi.org/10.1016/j.compedu.2007.09.017

Bloom, B. S., Englehart, M. D., Furst, F. J., Will, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Cognitive domain*. McKay.

Bonwell, C. C., & Eison, J. A. (1991). *Active learning: Creating excitement in the classroom* (ASHE-ERIC Higher Education Report).

Castelli, F., & Sarvary, M. (2021). Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Ecology and Evolution*, *11*(8), 3565–3576. https://doi.org/10.1002/ece3.7123

Chen, Y., & Roldan, M. (2021). Digital innovation during COVID-19: Transforming challenges to opportunities. *Communications of the Association for Information Systems*, *48*, 15–25. https://doi.org/10.17705/1CAIS.04803

Connolly, A., & Mutchler, L. (2021). A course plan for principles of IS programming to withstand COVID-19. *Communications of the Association for Information Systems*, *48*, 1–7. https://doi.org/10.17705/1CAIS.04801

Dick, G. (2021). Teaching online: Creating student engagement. *Communications of the Association for Information Systems*, *48*, 65–72. https://doi.org/10.17705/1CAIS.04809

Figl, K. (2010). A systematic review of developing team competencies in information systems education. *Journal of Information Systems Education*, *21*(3), 323–337.

Frost, R. (2021). Stealth theory through instructional scaffolding in the COVID-19 era and beyond. *Communications of the Association for Information Systems*, *48*, 26–34. https://doi.org/10.17705/1CAIS.04804

Gomez, E. A., Wu, D., & Passerini, K. (2009). Traditional, hybrid and online teamwork: Lessons from the field. *Communications of the Association for Information Systems*, *25*(33), 395–412. https://doi.org/10.17705/1CAIS.02533

Hvalshagen, M., Nittala, L., Raman, R., Sullivan, N., & Zolbanin, H. (2021). When worlds collide: Framing students' challenges with stay-at-home learning during COVID-19 through the lens of conflicting role identities. *Communications of the Association for Information Systems*, *48*, 227–235. https://doi.org/10.17705/1CAIS.04829

Kaufman, P., & Schipper, J. (2018). *Teaching with compassion: An educator's oath to teach from the heart*. Rowman and Littlefield.

Leidner, D. E., & Jarvenpaa, S. L. (1995). The use of information technology to enhance management school education: A theoretical view. *MIS Quarterly*, *19*(3), 265–292. https://doi.org/10.2307/249596

Lowell, A. L. (1926). The art of examination. *The Atlantic Monthly*, *137*(1), 58–66.

McGregor, J. (2021, March 30). The hybrid office is here to stay. The shift could be more disruptive than the move to all-remote work. *The Washington Post*. https://www.washingtonpost.com/business/2021/03/30/hybrid-office-remote-work-citigroup-ford-target/

Mitchell, A., Petter, S., & Harris, A. (2017). Learning by doing: Twenty successful active learning exercises for information systems courses. *Journal of Information Technology Education: Innovations in Practice*, *16*, 21–46. https://doi.org/10.28945/3643

Nilson, L. (2015). *Specifications grading*. Stylus Publishing.

Olsen, T. (2021). Using recent graduates as five-minute guest speakers to provide professional socialization and topical context for students. *Communications of the Association for Information Systems*, *48*, 109–114. https://doi.org/10.17705/1CAIS.04815

Osmani, M., Weerakkody, V., Hindi, N., Kapoor, K., Al-Esmail, R., & Eldabi, T. (2016). Skills and attributes of IT graduates: Evidence from employer's perspective. *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS)*, 1–9.

Sanandaji, A., & Ghanbartehrani, S. (2021). An evaluation of online flipped instruction methods during the COVID-19 pandemic. *Journal of Information Technology Case and Application Research*, *23*(1), 46–67. https://doi.org/10.1080/15228053.2021.1901360

Serva, M., & Fuller, M. (1999). The role of media use and active learning in higher education: The development of an instrument to determine the dimensions of teaching. *Proceedings of the International Conference on Information Systems*, 386–399. https://aisel.aisnet.org/icis1999/36

Sharadgah, T. A., & Sa'di, R. A. (2020). Preparedness of institutions of higher education for assessment in virtual learning environments during the COVID-19 lockdown: Evidence of bona fide challenges and pragmatic solutions. *Journal of Information Technology Education: Research*, *19*, 755–774. https://doi.org/10.28945/4615

Toney, S., Light, J., & Urbaczewski, A. (2021). Fighting Zoom fatigue: Keeping the Zoombies at bay. *Communications of the Association for Information Systems*, *48*, 40–46. https://doi.org/10.17705/1CAIS.04806

Wu, D., Bieber, M., & Hiltz, S. R. (2008). Engaging students with constructivist participatory examinations in asynchronous learning networks. *Journal of Information Systems Education*, *19*(3), 321–330.

Xie, X., Siau, K., & Nah, F. (2020). COVID-19 pandemic – online education in the new normal and the next normal. *Journal of Information Technology Case and Application Research*, *22*(3), 175–187. https://doi.org/10.1080/15228053.2020.1824884

DECISION SCIENCES INSTITUTE

Using Text Data Mining To Discover The NAICS Industry Code For A Business

Andres Fortino
New York University
New York, USA
Agf249@NYU.edu

Yongqi Tong
Sichuan
University,
Chendu, China
tongyongqi@stu.scu.edu.cn

ABSTRACT

This paper presents the development and testing of a text data mining tool to assist teachers, student entrepreneurs, and marketers to quickly find the NAICS code for a company using the description of the business. The tool compares specific parts of business plans with the official descriptions of NAICS code using text similarity scoring. We investigated using 4-digit and 6-digit NAICS codes and their standard descriptions. We found that using the 4-digit NAICS industry description produced a highly accurate match compared to the analysis conducted by a subject matter expert. Using the 6-code descriptions produced less reliable results. We investigated several text analytic tools for text analysis and found TF-IDF similarity scoring superior to other methods. Our research question was to investigate whether this similarity-scoring tool for students' text against industry descriptions would improve the search and comparison process to discover the best-fit code. We applied the tool to student's work from several sections of an entrepreneurship graduate class in a STEM master's degree. We investigated using serval parts selected from their business plans and pitches, including the entire plan, from 64 students' plans. We compared our predicted code match to their manually previously determined NAICS codes. We discovered that using detailed elevator pitches, executive summaries, and business descriptions produced the most accurate matches. On average, our matches exceeded 85% when compared to a subject matter expert code discovery. This result generated sufficient confidence in the text scoring system to integrate it as an easy-to-use and powerful tool in courses where industry classification is needed. It can also be used as a reference in early stage of entrepreneurship.

KEYWORDS: Text Similarity Score, Text Data Mining, TF-IDF, NAICS code, Industry Classification, STEM, entrepreneurship

INTRODUCTION

Business students or early-stage entrepreneurs will face a troublesome task when writing a business plan. They need to discover the correct NAICS code for their target company. This code is often the start of industry research to support the development of the various elements of defining a new business. The course instructor or relevant reviewer needs to determine whether the code is accurate based on the student's description of the bsuiness. The current way to determine the best NAICS code to match to a startup or a marketing endeavor is to perform a manual search of codes and rely on the best judgment of fit. Using the official NAICS website code search web interface, the user first determines the 2-digit codes that categorizes the business in general. They then select the appropriate 3-digit code and continue working their way down the tree to get to the more granular 6-digit classification of the company. This is cumbersome and often very time-consuming, and it is possible to choose the wrong industry code because of a mistake. It is further challenging in the case of students with little industry experience. If the student's business plan is not fully developed and the description is not clear, we may judge the student's assignment as not well completed. Faced with the need to correct and coach many students in this process, having a tool to get the student closer to the correct classification would be extremely useful.

Our work is based on text data mining techniques to help users quickly and accurately determine the proper NAICS code. We have developed a Python-based text analysis tool to score the text of the companies' business description against the NAICS industry descriptions We used it to check the work previously done by students in an entrepreneurship course. We found that this scoring tool is an excellent way to determine the specific industry for a given company. When we compared using this tool and not using this tool to determine the industry code, we found that this tool can quickly and easily obtain accurate NAICS codes. This tool ranked the similarity score of a company's text description based on the NAICS official website's industry description. We used the 2017 NAICS codes, using both the short code description with the longer and fuller descriptions appended to the shorter description. This tool can be used as the first-pass filter for determining NAICS codes, quickly obtaining the top ten most relevant from the 4-digit NAICS code lists. The codes are returned in ranked order of text similarity for the user to specifically select one or more codes that are more suitable for the company based on the output results and the industry description. The user then needs to choose from a 6-digit, more detailed list of codes for a relevant, more granular result. The use of this tool can save a great deal of time to find the proper NAICS code. Outputting the top ten codes can also better solve the situation that cross-industry companies may fit multiple NAICS codes. The user only needs to choose a few codes that are closest to the company. It can be seen from the industry description given on the official website that a four-digit code corresponds to a small number of six-digit codes. In other words, if people need to get a more specific code, they only need to get the result from several six-digit codes corresponding to the selected four-digit code. The tool dramatically shortens the query time while ensuring reasonable success.

LITERATURE REVIEW

To the best of our knowledge, this may be the first attempt to automatically identify the companies' NAICS code with their business plans or descriptions. However, text data mining in classification has been under study for quite some time, especially in the financial and economic fields. We emulated the work of Fortino et al. (Fortino, 2019), where they successfully applied text mining methods to choose teaching staff based on their resume. Also, the work of Chen HW. (Chen, 2010), in his economic data mining project, he developed a tool to analyze and classify different opinions on the websites. His work is an early method to process texts in economic analysis.

Kumar and his colleagues (Kumar, 2018) also developed an algorithm for unstructured mining data to perform a SWOT analysis, a widely used corporate strategy analysis methodology. Using TF-IDF to manipulate the cross-border e-commerce commodity's text corpus for getting the weight matrix of their names was developed by Luo (Luo, 2019). The vector that combined a weight vector and a word vector was put into LSTM to get HSCode. However, unlike Kumar, they did not emphasize unstructured data, which is more common.

The application of text data mining to classification follows two significant approaches. The first and very successful approach is to extract key features from the text to match the descriptions. Many researchers used this approach. For instance, Onan et al. (Onan, 2016) examined the prediction performance of five statistical keyword extraction methods based on five widely used ensemble methods. They pointed out that combining text-based learning with ensemble learning can improve the prediction performance and scalability of text classification schemes.

The other approach is to process the text analytic results using alternative machine learning algorithms — clustering, sentiment analysis, or ontological studies based on similarity scores. For instance, the vector space model of distributed representation was used for text clustering in Qimin's work (Qimin, 2015). In the preprocessing, he identified the discontinuous phrases of proper nouns and then represented the features as vectors, making it easy to calculate the distance between features. After clustering the feature, he used the co-occurrence matrix of text feature clusters to represent the document. This method can obtain more compelling features, reduce the size of features, and reduce model processing complexity. In Singh's work (Singh, 2019), he put forward a new SentiVerb system framework and take abbreviation extension as a part of spelling correction. He also put forward a new concept of affirmative words and introduce a negative threshold parameter. More recently, ontological studies have been widely used in the text data mining field. Witte R. and his colleagues (Witte, 2007) proposed a text mining method to analyze software documents at the semantic level using ontological studies.

Given that an enormous amount of online data has been generated every day, text data mining skills with online sources are becoming increasingly significant. Narasimha V. and his colleagues (Narasimha, 2020) clarified how to extract data from the Internet and track topic using text data mining skills, which helps readers obtain prominent opinions as quickly as possible. Similarly, Dhar A. (Dhar, 2018) developed a TF-IDF tool with a dimensionality reduction technique to categorize online documents. And the experiment shows that what they developed can achieve good results in high Viterbi collection and relatively noisy document feature vectors.

Moreover, Naïve-Bayes, a well-known algorithm, is also efficient and effective. Jiang (Jiang, 2006) successfully applied it to feature selection and achieved significant improvement in classification accuracy, which surprisingly improved the probability estimation. Also, Li (2016) did a more in-depth experiment. They classified news on the Internet and finished sentiment analysis using implementations of TF-IDF, Naïve-Bayes, and SVM. As it turns out, Naïve Bayes has its own merits when it comes to classification in some cases.

In summary, much of this work differs from our efforts in that we apply the text data mining methods and similarity scoring skills to classify one specific company's industry. Few workers, or even no workers to our knowledge, have tackled the automation of the identification of NAICS code for a company or marketing plan with some textual descriptions of the company. The work presented in this paper is simple: the use of the similarity scoring as an imperfect indicator could be used as a filtering tool to reduce the task from manually searching over 2000 NAICS codes to just the top 10, most likely as a first pass. In the analysis

This paper reports using various parts of business plans as the input data and the official industry description as the descriptions.

RESEARCH QUESTION

Can this text-similarity scoring tool that ranks the relevant text of a company's description useful for finding valid NAICS code?

Hypothesis

Relating to the primary objective of this research to test the efficacy of a literature search tool that improves the process of finding relevant literature for a literature review for research purposes by a novice, two hypotheses were proposed:

> *Hypothesis 1: Using the text-similarity score to rank potential industries for a company provides a significant improvement in the search process.*
>
> *Hypothesis 2: The text similarity score is an adequate indicator of the NAICS industry code's proper choice.*

METHODS

Text Data Mining

Many algorithms may successfully be used for text data mining, including TF-IDF, Naïve-Bayes, Support Vector Machines, neural networks, among many. Hence, we need to choose a suitable algorithm to build a model for industry classification. This article chooses two methods, TFIDF and Naïve-Bayes, to compare. These two algorithms can be used separately or combined.

<u>TF-IDF</u>

TF-IDF method is a process that converts documents into a numeric matrix with the combination of two terms: term frequency (TF) which means the frequency of a specific term t in document d, and inverse document frequency (IDF), generally estimating the significant role of a particular term t in a text document d.

TF is defined as the number of times a term appears in a document. Therefore, the greater TF means, the more important the word is in that document. TF can be calculated by the following algorithm, where N_t means how many times the term t showed in document d and

$\sum_{t=0}^{n} N_t$ is the total number of words.

$$TF_t = \frac{N_t}{\sum_{t=0}^{n} N_t}$$

IDF is measured by the following formula, where N is the total number of text documents and DF_t means the document frequency, which is the number of documents where this term t occurs.

$$IDF_t = \log \frac{N}{DF_t}$$

Now, TF-IDF methods for a term t in document d are computed using the formula stated below.

$$\mathrm{TF} - \mathrm{IDF} = \mathrm{TF} * \mathrm{IDF}$$

Naïve-Bayes

Naïve-Bayes Classification is a method based on Bayes theorem and assuming that the feature conditions are independent. That is:

$$P(a1, a1, \dots, an) = \prod_{i=1}^{n} P(a_i|c)$$

Naive Bayes is defined as follows.

$$g(e) = \arg max_{c \in C} P(c) \prod_{i=1}^{n} P(a_i|c)$$

where e = (a_1, · · · , a_n) is an instance and *g(e)* is the class assigned by Naïve-Bayes.

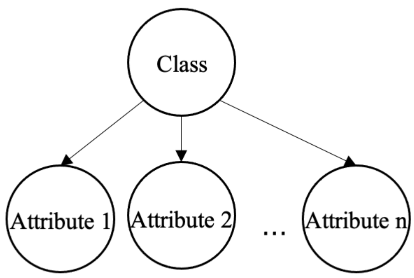


Figure 1 - This figure shows the structure of Naïve-Bayes graphically. In Naïve-Bayes, each attribute node has the class node as its parent but does not have any parent from attribute nodes.

Vector Space Model

The vector space model (VSM) is a typical algebraic representation of "text document" commonly used in information retrieval. This "bag of words" method can calculate the occurrence times of each word in the text and can be regarded as a simple and powerful representation. The vector space of a text set is constructed by representing each document as a vector containing the frequency of words or terms encountered in the document. Together, these document vectors represent a document by document matrix for the entire text. The

correlation of documents can be derived from these vectors, such as calculating the angle between document vectors by cosine measurement.
Similarity Scoring

After transforming the NAICS code description data into a latent space of lower dimensionality, the next step is to determine the similarity between the NAICS code description and a company description, such as the elevator pitch, part of the concept description, as well as two-way combination, etc. Each description is converted to a query vector in the same two-dimensional semantic space that we chose to perform the SVD for the NAICS code text data matrix. Then we used cosine similarity to measure the distance between a given query and the NAICS code description vector. We ranked the similarity scores in descending order. We can think of the highest-ranked code as being most similar to the enterprise being tested.

Implementation Details

The TF-IDF and Naive Bayes algorithms were implemented in Python code to use in experimenting with selecting between algorithms. The libraries we used are as follow:

- Pandas for reading and writing CSV files
- Sklearn and NLTK (Natural Language Toolkit) for preprocessing the text data
- CountVectorizer from sklearn.feature_extraction.text to calculate the frequency of words
- TfidfVectorizer from sklearn.feature_extraction.text to get the TF matrix
- MultinomialNB from sklearn.naive_bayes to implement naive bayes
- Cosine_similarity from sklearn to calculate job-course similarity
- cross_val_score from sklearn.model_selection for achieving k-fold validation
- Matplotlib for drawing

Experiment between Naive Bayes vs. TF-IDF algorithms

We compare TF-IDF and Naive Bayes algorithm with the advertising word part in the business plan of the graduate course report. This part is relatively concise and can well size up the characteristics of the company. We used two algorithms: TF-IDF and Naïve-Bayes algorithm, to compare the advertising words reported by the Structured Planning for Innovative Entrepreneurial Ventures course with each job description. We used these two algorithms to calculate the similarity score (cosine similarity) between a given job description and a program.

It should be noted that, after our initial experiments, it is difficult for both algorithms to directly obtain the 6-digit NAICS code. This is not only related to the fact that the industry description text given on the official website of each NAICS code is too sparse to support text similarity analysis. Through observation, we can understand that the companies with the same first four digits in their code belong to the same industry. It is too difficult to distinguish the small branches of companies in specific industries. Therefore, this article only calculates for the first four-digit list and combines all the industry descriptions of the NAICS code with the same first four digits as the model's training set. The result obtained in this way is also the top four NAICS codes. But this is too limiting because each 4-digit NAICS code only corresponds to about 3, 6-digit NAICS codes. Even if you desire to obtain the more accurate 6-digit NAICS code, you only need to start by distinguishing it simply from the 4-digit code and manually search through a smaller set of corresponding 6-digit code choices.

The prediction is based on the similarity score. For example, suppose the Naïve-Bayes algorithm calculates a given piece of text. In that case, the industry description scores for the two NAICS codes of 5322 and 6241 are 0.38 and 0.92, respectively, then the NAICS code of the company will be predicted as 6241 similar jobs. To avoid the situation where both similarity scores are too low to be used as a similarity program to work, we set a threshold defined as the 10th place among all similarity scores of the company's NAICS code matches. If specific similarity scores of courses rank below this threshold, we believe that the business probably does not belong in that industry. This will help filter out more irrelevant NAICS code, significantly save the time of manual selection, and meet the enterprise's diversified development requirements. A company may be involved in various industries so that it may match multiple NAICS codes. Choosing the top ten NAICS codes can cover most diversified companies' demands to find multiple NAICS codes.

Company Name	NAICS code	Similarity	Ranking	Output	Result
Info Inc	5614	0.083835	1	Yes	5416
	5416	0.083683	2	Yes	
	5619	0.08177	3	Yes	
	5611	0.074899	4	Yes	
	5182	0.072262	5	Yes	
	5511	0.07034	6	Yes	
	4431	0.064996	7	Yes	
	4483	0.062667	8	Yes	
	5191	0.060451	9	Yes	
	5617	0.05836	10	Yes	
	4412	0.058123	11	No	
	4883	0.057505	12	No	
	4911	0.057418	13	No	
	5419	0.05688	14	No	
	8112	0.055162	15	No	
	…	…	…	No	

Table 1 - Examples of encoding of algorithm scoring of a NAICS code description

To test the prediction's accuracy, we collected the student-determined NAICS codes described in the student's business plan. Since that was corrected and approved by an experienced faculty member, we then accepted that code accurately and backed it by a subject matter expert's finding. We then processed the business plans as we attempted to match the SME-approved code with our tool. We obtained the NAICS code for each student company. The final result will serve as our basis for judging the accuracy of our forecasted code. Business plans from 79 students attending the same entrepreneurship course taught by the same faculty member from 2015 to 2018. Those business plans that did not contain NAICS codes or well-defined business descriptions or detailed enough elevator pitches were deleted, resulting in a data set containing 64 business plans. We compared the prediction results of the two algorithms with the students' prediction results and approved by the SMEs, and calculated the prediction accuracy (the percentage of total patches between experts and algorithms). A summary of the SME approved codes to the predicted results, comparing the same data set's accuracy using Naïve-Bayes and TF-IDF follows.

Method	The number of accurately predicted samples	Accuracy
Naive Bayes	25	0.3906
TF-IDF	52	0.8125

Table 2 - Comparison of the prediction accuracy of the two methods

We can conclude that TF-IDF is a superior algorithm model for predicting a NAICS code for the data set collected through this comparative experiment.

APPLICATION TO PREDICT NAICS CODE

Entrepreneurs and business students need to determine a NAICS code in short order when starting a new company or to write a business plan. Using data mining, we can use tools to support this feature quickly. Using text data mining, we can score the full NAICS code industry description based on the text given by students or entrepreneurs, and provide job seekers with a list of the top ten industries to help them according to their own needs and the diversified development of the company determine the specific one or more NAICS code.

We downloaded a standard description of industry represented by every NAICS code from the North American Industry Classification System database. We performed TF-IDF similarity scoring of the text of each industry description and saved the model. Then we input the elevator pitch in the business plan completed by graduate students enrolled in a course Structured Planning for Innovative Entrepreneurial Ventures course into this model. We found that the effect is not good enough. The following table shows the similarity calculation results of then Cyberfence company, which is a cyber-security firm that provides security and risk assessment services and financial services. She pointed out the company's NAICS code is 541699.

Rank	NAICS code	Similarity score
1	5619	0.10070
2	561999	0.09661
3	541611	0.09430
4	5614	0.08770
5	333519	0.08726
6	561219	0.08583
7	561	0.08234
8	561329	0.08216
9	523	0.08109
10	812	0.08093
…	…	…
46	541699	0.05781

Table 3 - The result for the Cyberfence company

As we can see, the industry description of NAICS code 541699 is not very similar to the company. In the official description document of the NAICS code, we found that 541699 means this industry is engaged in providing advice and assistance to businesses and other organizations on scientific and technical issues (except environmental). Meanwhile, 541611 states that this industry comprises establishments that provide operating advice and assistance on management issues. It is not difficult to find that the two NAICS codes themselves are very similar, and both are more in line with this company's business. However, 541611 appeared in 3rd place, and 541699 was in 46th place.

By analyzing the six digits of the NAICS code, we can understand that the first two of the six digits represent specific and large-scale industries. For example, 11 is for Agriculture, Forestry, Fishing, and Hunting while 52 stands for Finance and Insurance. The third and fourth places determine a specific industry such as 5111 is for Newspaper, Periodical, Book, and Directory Publishers, and 5112 refers to Software Publishers. The last two digits determine the specific subdivision of the company in the industry. For instance, companies of NAICS code 512110 primarily engaged in Motion Picture and Video Production. Therefore, if we want to get a company's NAICS code more accurately, we can merge the same top 4 industry descriptions into one category and use the merged four NAICS code descriptions as the basis for the new model classification. After obtaining the top ten accurate NAICS codes, the users only need to choose the specific six being determined according to the company's main business type from a few NAICS codes.

Based on the model trained in the above steps, we tested 64 pieces of an elevator pitch in students' corporate planning books. We found that the most accurate predicted code for most companies is in the first 1-2 digits of the output result, and only a few companies have very large deviations. This may be related to the student's expressing themselves in the elevator pitch document. They may not clearly state the main business content of the company in the text. And the accuracy for elevator pitches is 0.764705882. However, we think this is not accurate enough. Then, we try to test different parts of the business plan to get a more accurate result.

We noticed that the business plan's executive summary contains a concise and clear introduction to the company. However, since the executive summary often includes financial situation which is not directly related to enterprise classification. Unfortunately, in the experiment, we found that this part will significantly interfere with the classification results. With this part of the business plan and the source text, it is more likely to classify companies into finance-related industries. As a result, when we remove this part from the executive summary and input the other parts into our model and calculate according to the above method, the accuracy is 0.823529412. Finally, we also combine it with the elevator pitch part as input for calculation, and the accuracy is as high as 0.852941176. The following table shows the results of our attempts.

Methods	Accuracy
Elevator Pitch	0.7647
Executive Summary without Financial Situation	0.8235
A combination of the above two	0.8529

Table 4 The result of three methods of input.

Based on the above results, we can conclude that our model can be well applied to the preliminary selection of the NAICS code. The user can input a text describing the company's business scope, which can be an elevator pitch, an executive summary without a financial situation, and a combination of the above two. And we can get accurate results through the output of the model.

CONCLUSIONS

Text mining has proved to be a valuable tool for us in the early stage of entrepreneurship. We have successfully applied TF-IDF to obtain a correct NAICS code. Our method can be used to guide students to complete the relevant requirements of the learning course efficiently. It can also help the entrepreneurial team to achieve the target results quickly.

FUTURE WORK

At present, we can only get the 4-digit NAICS code relatively accurately with the code we developed so far. The current data set is not enough to support accurately obtaining the 6-digit NAICS code. We hope to continue to refine our algorithm to get a 6-digit NAICS code with a higher degree of confidence in the future. We are also working to create a user-friendly interface to the Python code to allow the tool's use by a wider audience.

REFERENCES

Chen HW., Lee KR., Huang HH., Kuo YH. (2010) Unsupervised Subjectivity-Lexicon Generation Based on Vector Space Model for Multi-Dimensional Opinion Analysis in Blogosphere. In: Huang DS., Zhao Z., Bevilacqua V., Figueroa J.C. (eds) Advanced Intelligent Computing Theories and Applications. ICIC 2010. Lecture Notes in Computer Science, vol 6215. Springer, Berlin, Heidelberg.

Dhar A., Dash N.S., Roy K. (2018) Application of TF-IDF Feature for Categorizing Documents of Online Bangla Web Text Corpus. In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore.

Dhar A., Dash N.S., Roy K. (2018) Application of TF-IDF Feature for Categorizing Documents of Online Bangla Web Text Corpus. In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore.

Fortino, A., Zhong, Q., Huang, W., Lowrance, R. Application of Text Data Mining To STEM Curriculum Selection and Development, IEEE ISEC'19 Conference, Princeton University, NJ, March, 2019.

Jiang L., Zhang H. (2006) Learning Naive Bayes for Probability Estimation by Feature Selection. In: Lamontagne L., Marchand M. (eds) Advances in Artificial Intelligence. Canadian AI 2006. Lecture Notes in Computer Science, vol 4013. Springer, Berlin, Heidelberg.

Kumar, A., Dabas, V. & Hooda, P. Text classification algorithms for mining unstructured data: a SWOT analysis. International Journal of Information Technology (2018).

Li, J., Fong, S., Zhuang, Y. et al. Hierarchical classification in text mining for sentiment analysis of online news. Soft Comput 20, 3411–3420 (2016).

Luo, Y., Ma, J. & Li, C. Entity name recognition of cross-border e-commerce commodity titles based on TWs-LSTM. Electron Commer Res 20, 405–426 (2020).

Narasimha V., Sujatha (2020) An Efficient Classification Technique for Text Mining and Applications. In: Satapathy S., Raju K., Shyamala K., Krishna D., Favorskaya M. (eds) Advances in Decision Sciences, Image Processing, Security and Computer Vision. Learning and Analytics in Intelligent Systems, vol 4. Springer, Cham.

Onan, A., Korukoğlu, S., & Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. Expert Systems with Applications, 57, 232–247.

Qimin, C., Qiao, G., Yongliang, W. et al. Text clustering using VSM with feature clusters. Neural Comput & Applic 26, 995–1003 (2015).

Singh, S.K., Sachan, M.K. SentiVerb system: classification of social media text using sentiment analysis. Multimed Tools Appl 78, 32109–32136 (2019).

Witte R., Li Q., Zhang Y., Rilling J. (2007) Ontological Text Mining of Software Documents. In: Kedad Z., Lammari N., Métais E., Meziane F., Rezgui Y. (eds) Natural Language Processing and Information Systems. NLDB 2007. Lecture Notes in Computer Science, vol 4592. Springer, Berlin, Heidelberg.

Witte R., Li Q., Zhang Y., Rilling J. (2007) Ontological Text Mining of Software Documents. In: Kedad Z., Lammari N., Métais E., Meziane F., Rezgui Y. (eds) Natural Language Processing and Information Systems. NLDB 2007. Lecture Notes in Computer Science, vol 4592. Springer, Berlin, Heidelberg.

DECISION SCIENCES INSTITUTE

Simple Interaction Finding Technique (SIFT) – A Simple Methodology to Generate Novel Hypotheses from Complex Datasets

Murtaza Nasir
University of Massachusetts Lowell, 72 University Ave, Lowell, MA 01854,
academic@murtaza.cc, 605-202-1804

Nichalin Summerfield
University of Massachusetts Lowell, 72 University Ave, Lowell, MA 01854,
nichalin_summerfield@uml.edu, (978) 934-5525

Serhat Simsek
Montclair State University, 1 Normal Ave, Montclair, NJ 07043,
simseks@montclair.edu, (973) 655-4000

Asil Oztekin
University of Massachusetts Lowell, 72 University Ave, Lowell, MA 01854,
asil_oztekin@uml.edu, (978) 934-2714

ABSTRACT

Machine Learning (ML) models have become ubiquitous in all spheres of research and decision-making. Understanding ML models as well as the data-generation-process (DGP) for the dataset under examination are important. Most highly accurate ML models are blackboxes that aren't interpretable. In this work, we propose a methodology that can help elicit important information from any ML models. Our methodology allows the use of any highly-accurate ML model to find interactions between variables in the dataset. This can allow for a better understanding of the underlying DGP by using a data-and-model agnostic process to synthesize new knowledge about the underlying phenomenon.

KEYWORDS: Interpretable machine learning, interactions, methodology development

DECISION SCIENCES INSTITUTE

Demystifying the Relationship between the Corporate Financial Performance and ESG Disclosures

Nitin Jain
PhD Student, IIM Udaipur, India
nitin.2020phd@iimu.ac.in

ABSTRACT

There has been a growing concern around firms' need to place more thrust on environmental, social, and governance (ESG) initiatives. The extant research does not seem to be conclusive enough. Some findings assert a positive, a negative, and some say it is a u- or inverted u-shaped relationship between the ESG and financial metrics. Therefore, it is vital to analyze this relationship more deeply. This study investigates the relationship between corporate efficiency and sustainability to understand if high ESG scores are associated with a higher profitability. It adds to the research on ESG by conducting an empirical analysis on the S&P 500 firms for a ten-year time-period from 2011-2020. It sheds light on the concave nature of the relationship between ESG and financial performance and how issues such as endogeneity arise and could be mitigated.

KEYWORDS: ESG Disclosure, Sustainability, Transparency, Financial Performance, Stakeholder Theory, and Shareholder Theory

INTRODUCTION

The link between sustainability and the financial performance of a firm remains an open question since 1987 when the Brundtland Commission defined sustainable development. Since then there has been a growing trend in the ecosystem of corporate sustainability, from voluntary engagement in sustainable activities to de facto needs owing to both regulatory pressure and social expectations (Brockett & Rezaee, 2012). The financial scandals such as Enron, Worldcom, and the global financial crisis (2007-2009) paved way for more sensitivity towards disclosures around ethical values, and long-term sustainability performance (Rezaee & Tuo, 2019). Sustainability is now considered more comprehensive than corporate social responsibility (CSR) (Soytas, Denizel, & Usar, 2019). The business models and management have evolved fundamentally due to the increase in the number of firms embracing sustainability strategies and disclosing information related to ESG initiatives (Xie, Nozawa, Yagi, Fujii, & Managi, 2019).

Be it individual or institutional investors, governments, buyers, and suppliers; there is a growing interest in environmental, social, and governance (ESG) - related issues (Escrig-Olmedo, Muñoz-Torres, & Fernández-Izquierdo, 2013; Hill, Ainscough, Shank, & Manullang, 2007). There is a belief that investors focus more on non-financial disclosures as compared to financial ones (Cohen, Holder-Webb, Nath, & Wood, 2011). The environmental initiatives include aspects such as greenhouse emissions, hazardous wastes, and environmental penalties. In contrast, the social dimension considers metrics such as community spending, and percentage of employee turnover, etc. While the governance dimension focuses on attributes such as board duration (Yu, Guo, & Luu, 2018). Hence, in the context of the sustainability performance of organizations, ESG scores appear to be appropriate measures of evaluation (Ahi, Searcy, & Jaber, 2018) and have a positive connection to a firm's sustainability (Rajesh & Rajendran, 2020). According to the Governance and Accountability Institute (2017), the percentage of S&P500 companies embracing sustainability reporting increased from about 20 percent in 2011 to 72 percent in 2013 and 85 percent in 2017. Rating agencies such as MSCI, Thompson Reuters Asset4, and financial information providers such as Bloomberg have started reporting ESG data. Bloomberg, for example, provides both a combined ESG score and individual environmental, social, and governance scores for firms.

Across the different echelons of the supply chain, multiple entities consume energy and contribute to issues across the social, environmental, and economic dimensions by their sub-sustainable practices such as waste disposal (Mittal & Sangwan, 2014). Therefore, along with the product value, there are environmental and social burdens that tag along at the different production process stages (Seuring & Müller, 2008). Concerns have been raised on the manufacturing firms' lack of adherence to environment-friendly policies, and indulgence in over-exploitation of natural resources, besides generating vast amounts of toxic wastes (Villena & Gioia, 2020). However, the focal firms, i.e., the big multinationals, are held responsible for their suppliers' social and environmental negligence (Seuring & Müller, 2008). Renowned brands have been blamed for poor working conditions (Graafland, 2002) or impact on the environment at their suppliers' end (Seuring, 2001). Energy efficiency-related concerns have also been raised in recent years. According to the International Energy Agency (IEA), energy efficiency can boost economic growth and prevent greenhouse gas emissions, but the global rate of progress is a concern (IEA, 2019b). For the past three years, the rate of global primary energy intensity improvement, which indicates how heavily energy is used by global economic activity, has been below the 3 percent threshold needed for achieving global climate and energy goals. The 3 percent threshold could generate an additional USD 2.6 billion of economic output – nearly the French economy's size (IEA, 2019a; IEA, 2019b).

Measuring the supply chain's sustainability performance, therefore, has a far-reaching effect on the firms' profitability and achieving future competitiveness (Nikolaou, Tsalis, & Evangelinos, 2019). From an investor's perspective, socially responsible firms create value while minimizing risk. While from a consumer's view, firms that act responsibly deliver goods and services which protect the milieu, satisfy needs, protect consumers, and are reasonably priced (Minutolo, Kristjanpoller, & Stakeley, 2019). Earlier the focus of the research had been financial disclosures, it has now evolved into non-financial disclosures due to the availability of data (Rezaee & Tuo, 2019). In the past two decades, a large amount of empirical research has investigated the relationship between corporate sustainability and corporate financial performance (CFP) to understand the implications of stakeholder management. This investigation is important for the industry as they have invested significantly in social and environmental practices in the recent few years (Wang & Sarkis, 2013). Two key views prevail in the extant literature – the Porter hypothesis and Agency problems and inefficient resource allocation. According to the Porter hypothesis (Porter & Van der Linde, 1995), corporate social responsibility (CSR), especially environmental initiatives, lead to innovation which pushes the revenue more than the cost incurred. Therefore, CSR and CFP are positively related. In contrast, some researchers (Friedman, 2007; Sternberg, 1997) assert that CSR activities add to costs due to agency issues and inefficient resource allocation, which render a firm unfavorable in a free market.

The past studies show a positive (Orlitzky, Schmidt, & Rynes, 2003), negative (Margolis & Walsh, 2003), and neutral (McWilliams & Siegel, 2001) relationship between CFP and corporate sustainability performance (CSP)/ESG disclosure scores. The uncertainty on the nature of the CFP-sustainability relationship could be due to multiple reasons – a wide range of CFP metrics (Xie et al., 2019), CSR is a multi-dimensional construct (Waddock & Graves, 1997), and functional misspecification by assuming a linear relationship as shown by Barnett and Salomon (2006). Studies such as Trumpp and Guenther (2015) exhibited a non-linear relationship. Therefore, differences in metrics and functional specification, i.e., linear versus curvilinear (McWilliams & Siegel, 2001) and lagged versus current variables (Hart & Ahuja, 1996), could also be key factors for getting mixed and insignificant relationships in the past research.

A question that comes to mind is, can high ESG disclosure scores lead to a monotonic improved financial performance? This is imperative given the lack of consensus in the existing research. Therefore, an opportunity area in the extant research is around the scope of the investigation and analyzing when and what could cause the relationship to exist. According to Seuring and Müller (2008), most of the research has leveraged a variety of self-reported measures, from a single source, and focuses on either the social or environmental aspects separately. This paper, therefore, examines how firms' ESG disclosures can impact their financial performance. It extends the extant literature by having much larger firm years in terms of balanced panel and leverages organizational theories for the theoretical motivation. It is organized as follows. First, there is a discussion on the relevant organizational perspectives. This would facilitate deducing the hypothesis. This is followed by a discussion of the data and methodology. Thereafter, the results are presented, and findings are discussed. Finally, it concludes with the key takeaways, limitations, and future direction.

LITERATURE REVIEW & HYPOTHESIS

Friedman's share-holder management theory endeavored to improve financial performance and maximize shareholders' benefits (Friedman, 2007). While those advocating the need for sustainable business assert that it promotes stakeholder management and thereby focuses on combating externalities and maximizing social value related to ESG issues (Xie et al., 2019). Several theories explain why firms disclose ESG information voluntarily (Xie et al., 2019). Our literature review revolves around these theories and ESG-related initiatives and disclosures.

Organizational Theories

Legitimacy Theory

Legitimacy is the generalized perception that an entity's actions are appropriate within some socially constructed beliefs, values, and norms (Suchman, 1995). It provides a powerful mechanism to understand voluntary social and environmental reporting by firms (Tilling, 2004). According to the legitimacy perspective, the right to exist is an exogenous grant provided by a social contract to the firm and should be renewed continuously (Minutolo et al., 2019). The survival of a firm is threatened if the society feels it has not fulfilled its side of the social contract (Lokuwaduge, Chitra Sriyani De Silva & Heenetigala, 2017). Therefore, the firm's intent to represent a moral claim to the contract is exhibited by the ESG activities (Scherer & Palazzo, 2011). An entity gains legitimacy when its value system in sync with that of the larger system in which it resides (Lindblom, 1994). Hence, the ESG disclosure aims to attain social legitimacy for environmental or social impacts caused due to a firm's operation. Thus, a firm's ESG disclosure is strongly driven by its ESG legitimacy (Lokuwaduge, Chitra Sriyani De Silva & Heenetigala, 2017). It is also a tool to improve corporate reputation (Brammer & Pavelin, 2008). ESG disclosure also mitigates information asymmetry between a firm and outsiders, say, investors, and thus can mitigate the lemon problem by exuding reliable signals to the market. Socially responsible investments (SRI) are a key testimony to this belief and are found to be relatively more preferred over conventional funds by socially-oriented investors (Riedl & Smeets, 2017).

Stakeholder Theory

A stakeholder is an individual or a group of people, who are affected by the activities of a firm or can impact the firm's activities in achieving its objectives (Freeman & McVea, 2001). Under the stakeholder theory, it is asserted that firms perform better when they manage the interests of their stakeholders well (Minutolo et al., 2019). As sustainability decision-making is linked to the stakeholders' needs, ESG reporting is, therefore, tied to the stakeholder theory (Freeman & McVea, 2001). There is also a notion of the residual risk's information effect. According to the proponents of this argument, the non-accounting parameters' ratings of a firm reveal how it manages the risks it faces. Hence, a higher ESG rating would point to a mean lower risk relative to the market (Kurtz, 2005). It is argued that each stakeholder would assess the ESG information differently. For example, consumers might focus on labor policies while investors could be keener on the residual risk. Also, different stakeholders could exert varying degrees of legitimacy, power, and urgency (Mitchell, Agle, & Wood, 1997).

Voluntary Disclosure Theory

As per the voluntary disclosure theory, a better sustainability performance tends to lead to more signals from firms about their sustainability and therefore, they have better financial performance (Rezaee & Tuo, 2019). While corporate information disclosure might be voluntary in most geographies, investors may discount the undisclosed information as unfavorable. Therefore, firms are in a dilemma and faced with a trade-off to what extent they should disclose information (Xie et al., 2019). Another view is around the ESG 'advertising' impact. So, under the lens of marketing, the cost and benefits associated with sustainability would be akin to those of an advertising campaign (Sahut & Pasquini-Descomps, 2015).

ESG Initiatives and Disclosures

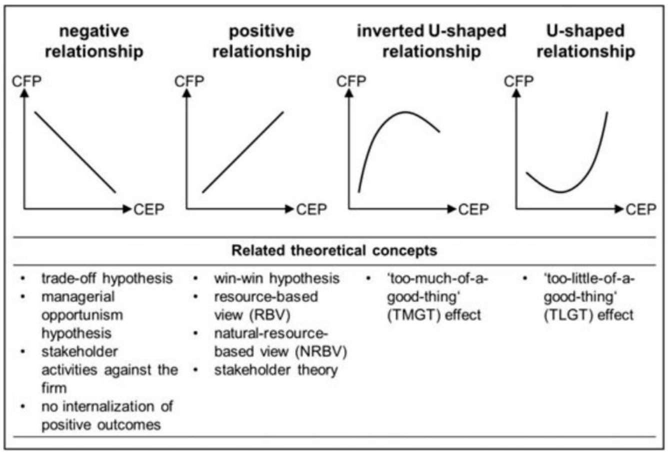
ESG practices are dissimilar to accounting practice so it provides additional information and ESG disclosure offers a better understanding of the firm's policies (Li, Gong, Zhang, & Koh, 2018). Nowadays, ESG reports encompass not only how firms deal with ESG topics, but also the steps they undertake to support global engagement and finding solutions (Ionescu, Firoiu, Pirvu, & Vilag, 2019). The ESG data encompasses a plethora of metrics such as measures of firms' environmental impact, human rights, and labor policies, and corporate governance measures (Romero, Jeffers, Lin, Aquilino, & DeGaetano, 2018). This is critical to understand as there is an asymmetry between the market and firms for a lot of socially responsible initiatives. Therefore, the ESG score is a signal to differentiate the quality of assertions on these initiatives (Simaens & Koster, 2013). It improves visibility and transparency in a firm's environmental, social, and governance factors (Dubbink, Graafland, & Van Liedekerke, 2008) and thus reduces agency costs (Li et al., 2018). Even conventional fund managers consider ESG attributes while making investment decisions (Van Duuren, Plantinga, & Scholtens, 2016).

Another reason why ESG initiatives should improve firm performance is due to "cost of capital" reduction. Thus, the costs involved in establishing a socially responsible structure within a firm is offset by the decrement in its cost of capital (Sahut & Pasquini-Descomps, 2015). Another perspective is provided by the "best practices" anticipation theory. According to this view, the costs arising from the implementation of a sustainable structure such as environmental regulations are offset by the improving productivity of the firm (Porter, 1991). The benefits happen in two ways Firstly, sustainable firms have a better distribution of costs, and hence, more stable cash flows. Secondly, they become leaders in embracing these best practices, are more forward-thinking, which leads to an increase in firm and shareholder's wealth (Sahut & Pasquini-Descomps, 2015).

The above may indicate the competitive edge ESG disclosures bring to a firm. However, prior studies on CFP and sustainability measures/ESG disclosures have been found to yield a mixed relationship. Studies found that a higher environmental disclosure could be translated to analysts' earnings forecasts. Governance disclosure benefitted the principal, so firms that follow strict disclosure norms would employ more capable managers and hence, have better financial performance (Hermalin & Weisbach, 2001). Table 1 summarizes the mixed relationship between CFP and ESG disclosures as found by extant studies. The mixed results could also point to confounding effects (Soytas et al., 2019). Amidst this background, it makes it pertinent to delve deeper into the relationship and explore a curvilinear relationship as most of the studies adopt a linear model specification. This would also throw light on the presence of a threshold ESG disclosure effort beyond which the CFP would increase or decrease monotonously and also help to understand the marginal impact of ESG disclosures on CFP.

There are certain activities that firms are compelled to do, such as, financial disclosures. But there are also discretionary initiatives which some firms do more than others, and this makes it interesting to investigate if there is an optimal point of indulging in these initiatives to maximize the CFP or if the relationship between ESG disclosures and CFP is strictly monotonous. Figure 1 illustrates the different possible relationships between corporate environmental performance (CEP) and CFP. Both the "too-little-of-a-good-thing"(TLGT) and "too-much-of-a-good-thing" (TMGT) assume a curvilinear relationship between CEP and CFP. The voluntary environmental-related initiatives are contrary to the firm's objective of profit maximization (King & Lenox, 2002). Therefore, the managerial opportunism hypothesis suggests that if the executive's compensation is linked to short-term shareholder value, then there is a negative relationship between CEP and shareholder value (Trumpp & Guenther, 2015). Secondly, increasing compliance costs for pollution reduction measures could also contribute to the negative relationship between CEP and CFP. However, the stakeholder perspective hinges on the legitimate interests of different entities – employees, buyers, suppliers, etc. Thus, by improving social/environmental aspects, a firm experiences an increase in costs but also a rise in brand equity.

Figure 1: Possible relationships between CEP and CFP



Source: Trumpp and Guenther (2015)

Table 1: Impact of Disclosures on Firm Performance

S.no.	Impact of ESG Disclosures	Sample Studies
1	ESG reporting improves stock market returns	Weber (2014)
2	Higher ESG transparency improves firm value	Yu et al. (2018)
3	Integrating ESG reporting creates a relatively better market valuation over stand-alone reporting	Mervelskemper and Streit (2017)
4	There is no relationship between environmental disclosure and CFP	Qiu, Shaukat, and Tharyan (2016)
5	There is a positive relationship between social disclosure and CFP	Qiu et al. (2016)
6	Social and economically sustainable supply chain practices are positively related to CFP	Wang and Sarkis (2013)
7	ESG performance improves firm performance, CEO power strengthens it further	Velte (2020)

We hypothesize the following:

H1: ESG disclosures have a concave (curvilinear) relationship with CFP

Thus, we contend that there is a critical point beyond which firms do not benefit from an increase in their ESG disclosure scores. This is important to understand how the ESG disclosures could motivate firms and to what extent would these improve financial performance.

DATA

The extant literature has used various sources and Bloomberg is also one of them as shown in Table 2. For our analysis, we construct a unique panel of 338 S&P 500 firms for the years 2011-2020, i.e., a balanced panel of 3380 firm years. We source the financial and ESG data fields from Bloomberg. Bloomberg collates the ESG information from annual reports, sustainability reports, press reports, and also includes board-related details such as women on board (Xie et al., 2019).

The ESG disclosure field ranges from 0.1 to 100, the higher, the better (Yu et al., 2018). ESG score can act as an indicator for the sustainability performance of organizations (Rajesh & Rajendran, 2020). As in the extant literature, we have defined the accounting performance of the firm in terms of its return on assets (ROA), which indicates the efficiency of the firm in using its asset. It is defined as the ratio of net profit and total assets. Tables 3 and 4 present the descriptive statistics and describe the variables, respectively. Table 5 gives the frequency distribution of firms by their GICS sector classification. Two-thirds of the analyzed firms belong to just four industry sectors – Information Technology, Industrials, Consumer Discretionary, and Health Care. Therefore, to capture the impact of the industry sectors, we introduced dummy variables for these in the model/consider their fixed effect.

Table 2: Data Sources in the Extant Research

S.no.	Data Source	Sample Studies
1	Bloomberg	Yu et al. (2018); Wang and Sarkis (2013); Tamim and Sebastianelli (2017); Minutolo et al. (2019); Li et al. (2018)
2	WRDS COMPUSTAT	Wang and Sarkis (2013); Soytas et al. (2019)
3	Thomson Reuters	Cek and Eyupoglu (2020); Rajesh and Rajendaran (2020); Sahut and Pasquini-Descomps (2015); Velte (2020)
4	CSRHUB	Soytas et al. (2019)
5	FACTIVA	Soytas et al. (2019)
6	MSCI KLD	Brogi and Lagasio (2019)

We also computed the variance inflation factor (VIF) and found that it was less than 4 for all the control variables if we include just ESG or ESG^2 but not both. If we include both ESG, ESG^2 in the VIF computation, then the VIFs of these two are 29. It is important to highlight here the ESG can take values between 0.1 and 100 and as shown in Table 3 and Figure 2, the majority of the ESG values seem to be concentrated between 10 and 60 due to which the VIF increases when we consider both the versions of ESG in the VIF computation. Another important observation is around outliers in the data. Table 3 on descriptive statistics shows some extreme minimum and maximum values. However, these kinds of extreme values are technically not outliers in the financial data as the ratios can fluctuate wildly due to idiosyncratic and systematic reasons. Also, the values between the 5th percentile and 95th percentile seem plausible. Without knowing the business reasons, it would not be proper to remove firms that may show extreme values or winsorize those values. Also, their removal could lead to a reduction in the sample size of the balanced panel or make the panel unbalanced. Therefore, for the current analysis, no observations have been removed because of statistical extreme values.

Figure 2: Average GICS sector-wise ESG scores from 2011-2020

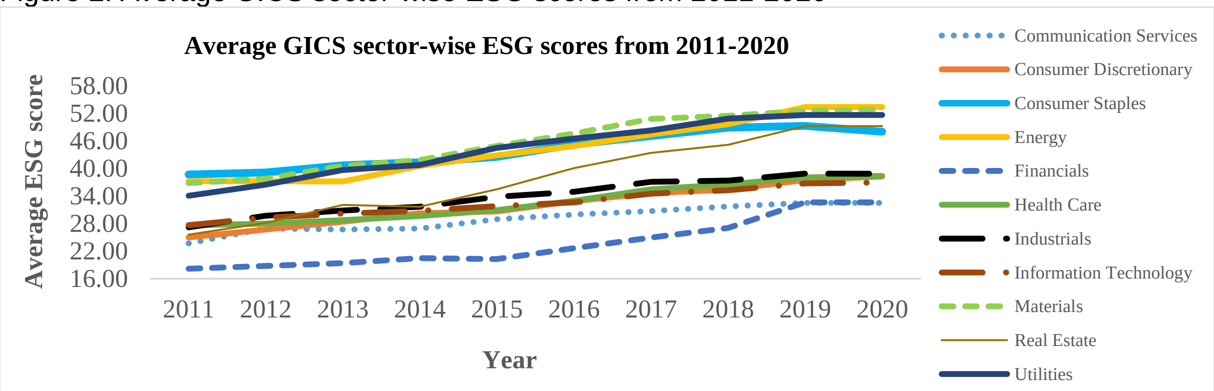


Table 3: Descriptive Statistics

Variable	5th percentile	95th percentile	Mean	Stdev	Min	Max
Debt_Assets	0.098	59.08	29.04	19.20	0.00	243.87
Log(Total Assets)	7.55	11.79	9.58	1.31	4.66	13.48
Cash_Ratio	0.034	2.83	0.78	1.02	0.00	9.84
Sales_Growth	-13.1	33.7	7.58	24.80	-79.83	883.91
ROA	-1.26	19.55	7.82	7.48	-48.13	48.45
ESG_lag1	12.8	59.24	33.96	15.54	9.92	78.01
ESG_lag1 Square	164.1	3510.1	1,394.90	1,144.26	98.35	6,085.30

Note: N = 3,380 observations (firm-years) with 10 observations for 338 firms from 2011 to 2020.

Table 4: Variables' Description

Label	Variable	Description
Debt_Assets	Debt to asset ratio	Leverage ratio that shows the ratio of assets funded by borrowing compared with the percentage of resources funded by investors.
Log(Total Assets)	Total Assets in log form	Total assets of a firm.
Cash_Ratio	Cash Ratio	Liquidity measure that reflects a firm's capacity to meet short-term obligations using cash, and cash equivalents.
Sales_Growth	Sales Growth	Percentage growth in sales over past fiscal.
ROA	Return on Assets	The profitability of a firm relative to its total assets.
ESG_lag1	ESG Score of past year	Environmental, Social and Governance scores computed by Bloomberg

Table 5: Frequency Distribution & Average ESG score by GICS Sector

GICS_Sector	Frequency	Percentage	Cumulative Percentage	Average ESG Score
Information Technology	62	18.34%	18.34%	29.07
Industrials	59	17.46%	35.80%	31.99
Consumer Discretionary	52	15.38%	51.18%	44.10
Health Care	51	15.09%	66.27%	44.39
Consumer Staples	28	8.28%	74.56%	23.72
Materials	22	6.51%	81.07%	32.65
Communication Services	16	4.73%	85.80%	34.05
Energy	16	4.73%	90.53%	32.60
Utilities	16	4.73%	95.27%	45.73
Financials	11	3.25%	98.52%	38.02
Real Estate	5	1.48%	100.00%	44.46

Note: There are 338 unique firms and each of these is associated with a GICS sector
The average ESG score for each sector is computed for the period 2011-2020

METHODOLOGY

The CFP at time t is a function of the cost (accounted by $-\beta_2^s$) and benefits (accounted by + β_1^r) of the ESG disclosures. Therefore, it can be represented as below, assuming r > 0 and s > 0.

$$CFP(t) = \beta_0 + \beta_1\{ESG(t-1)\}^r - \beta_2\{ESG(t-1)\}^s + control\ variables + \epsilon$$

If both r and s are equal to 1, then we have a perfectly linear relationship. However, if r < s, then we have an inverted U-shaped relationship, and for r > s, we get a U-shaped curvature. It is important to emphasize the need to explore a curvilinear relationship. This paper posits that there is an inverted U-shaped relationship between the CFP and ESG disclosure scores. Thus, we investigate the relationship via the fixed effects (fe) model as below.

$$\begin{aligned} ROA(i,t) = \beta_0 + \beta_1 ESG\_(i,t-1) + \beta_2\, ESG^2(i,t-1) + \beta_3 Log(Total\ Assets)(i,t) \\ + \beta_4 \frac{Debt}{Assets}(i,t) + \beta_5 Cash\ Ratio\ (i,t) + \beta_6 Sales\ growth\ (i,t) \\ + \beta_7 Year\ Fixed\ Effect + \beta_8\ GICS\ Industry\ Fixed\ Effect + \epsilon(i,t) \end{aligned}$$

The literature operationalizes financial performance by both accounting- (e.g. Return on Assets (ROA), Return on Equity (ROE)) and market-based measures (e.g. Tobin's Q). ROA and Tobin's Q also represent the long-term performance of a firm (Soytas et al., 2019). For our analysis, we consider ROA. Control variables and the industry, as well as year effects, have been captured in the research design to mitigate endogeneity concerns. For example, both the size of the firm and the number of employees are very relevant control variables since they can impact the relationship between ESG disclosures and CFP. The ratio of debt to assets has been considered as a proxy for unsystematic risk. Firms with a reasonably good performance on ESG are perceived to be less risky and hence, expected to have a lower cost of debt also (Orlitzky & Benjamin, 2001). Also, it is expected that larger firms would need to maintain their social contracts more actively and need to disclose better and signal to the market about their socially responsible behavior (Minutolo et al., 2019). Keeping in mind these intricacies, we have chosen the control variables for our analysis.

From the research design perspective, the discernment of the ESG-CFP relationship can be impacted by multiple reasons such as measurement issues, modeling shortcomings, and statistical challenges such as endogeneity (Soytas et al., 2019). It is critical to discuss endogeneity in this context. Firms may find it easier to make voluntary sustainability initiatives if they are more affluent and have slack resources. Therefore, performing an ordinary least regression would lead to biased estimates and not lead to the right causality estimates. Even when we try to mitigate the omitted variable bias (OVB) by controlling for firm-level productivity via factors such as R&D expenses, the estimates may not still be unbiased as some aspects of a firm's productivity cannot be observed (Olley & Pakes, 1996). We explore both the random effects (Re) and the fixed-effects (Fe) model for the analysis and choose between them based on Hausman's test.

A two-way Fe model for panel data is given by

$$Y(i,t) = \beta_0 + \mathrm{B}_1 X(i,t) + \mu(i) + \lambda(t) + \epsilon(i,t).$$

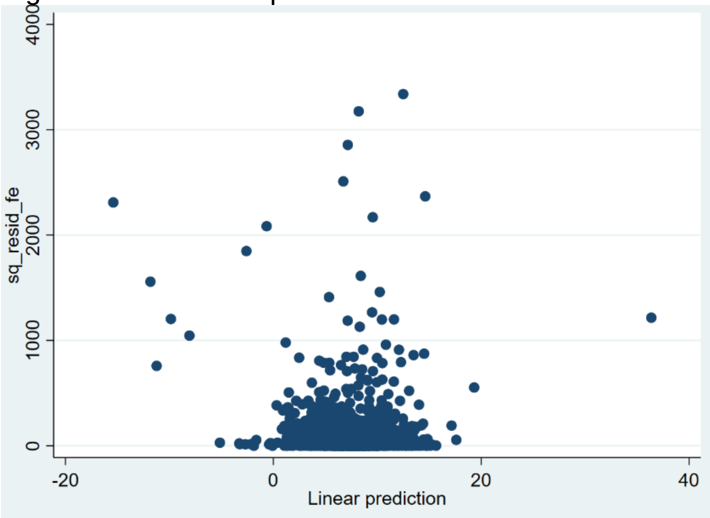
Here we assume that the unobserved heterogeneity across firm i may contribute to heterogeneity. Since $\mu(i)$ is unobserved we cannot control it. It is assumed to be correlated with X. In contrast, a Re model assumes that $\mu(i)$ are random and independent of X. To choose between them, we can perform Hausman's test in which the rejection of null leads to acceptance of the Fe model under the assumption of homoscedasticity. If $\mu(i)$ is independent of X, then Fe and Re are consistent but Re is more efficient. However, when $\mu(i)$ is correlated with X, then only Fe is consistent. Estimation of Fe models needs sufficient within-variation in X. Also, the observation within a panel could be correlated. Not accounting for these could lead to downward-biased standard errors.

RESULTS and DISCUSSION

In the linear version (models 1 and 2), we find that ESG(t-1) is not a significant variable. We also executed models 1 and 2 without robust errors so that we could conduct the Hausman test which rejected the null hypothesis that the preferred test is Re. Therefore, fe is the preferred model here. However, to confirm if the model specification is fine, we also plotted the residuals square against the estimated ROA from the fe model (Model 2) in Figure 3.

Figure 3 alludes to heteroscedasticity and/or misspecification due to the pattern of the residual squares. We, therefore, introduced a square term, ESG^2(t-1) too as an explanatory variable in the model. We also clustered the errors at the firm level to deal with both heteroscedasticity and auto-correlation as the robust clustered errors take into account these issues. Models 3 and 4 are the curvilinear versions with robust clustered errors. We also executed these two models without clustered errors to conduct the Hausman test, which again showed the fixed effect version to be the preferred model. Therefore, the recommended model is Model 4.

Figure 3: Residual Squares versus $\widehat{ROA}$ in linear Fe model (Model 2)



We find that in the linear versions – Models 1 and 2 – ESG (t-1) is insignificant. However, this insignificance is due to the incorrect functional form as seen in Models 3 and 4 in which both ESG(t-1) and ESG^2(t-1) become significant and indicate a concave relationship between ROA(t) and ESG(t-1) due to the positive coefficient for ESG(t-1) and the negative coefficient for ESG^2(t-1). After performing the Hausman test, the Fe model (Model 4) is found to be consistent in our case. Thus, the proposed model is:

$$\begin{aligned} ROA(i,t) = 2.6285 + \ 0.2283 ESG(i,t-1) - 0.0033\ ESG^2(i,t-1) \\ + 0.411 Log(Total\ Assets)(i,t) - 0.1042 \frac{Debt}{Assets}(i,t) \\ + 0.6973 Cash\ Ratio\ (i,t) + 0.0322 Sales\ growth\ (i,t) \\ + \beta_7 Year\ Fixed\ Effect + \ \beta_8\ GICS\ Industry\ Fixed\ Effect + \ \epsilon(i,t) \end{aligned}$$

Therefore, the critical point for ROA to start decreasing is 0.2283/2(0.0033) = 34.6. Therefore, the marginal effect of ESG is increasing as ESG increases from 0.1 to 34.6, and thereafter, it starts decreasing, keeping other factors constant. The other signages are also logical as higher debt to assets is prone to reduce profitability and both cash ratio and sales growth are expected to be positively related to ROA. The findings are very important as they allude to a curvilinear relationship between firm profitability and its ESG disclosure scores. Therefore, it is interesting to investigate if voluntary disclosures would reach a threshold level after which firms would not be motivated to pursue these initiatives. The current study is a good starting point in this direction.

It does have some limitations. The sample is of US-based firms only. So, the analysis needs to be conducted for a wider set of countries. Secondly, the analysis needs to be validated for different period balanced panels. For example, 2005-2020, 2015-2020 to see if the results hold. Heterogeneity needs to be investigated more thoroughly. A future step could be to consider an instrumental variable analysis, besides also replacing ROA with Tobin’s Q as the dependent variable.

CONCLUSION

The theme of this paper is a growing area of research as there is a lot of emphasis on ESG disclosures, sustainability initiatives across geographies. However, despite a fast-evolving body of research, the exact nature of the relationship between ESG disclosure scores and CFP is still uncertain. The theme of this paper is to investigate this relationship. This paper contributes in terms of its findings and larger sample size, besides handling endogeneity-related concerns. It finds a curvilinear relationship between ROA and first-order lag ESG scores, which contradicts most of the extant research. The findings are very informative as they allude to a threshold up to which firms may indulge in voluntary disclosures. Future directions include both improving the methodological rigor as well as considering a wider geographical scope to validate the findings. The current study is expected to be a good starting point to pursue research in the realm of sustainability in supply chains.

Table 5: Model Estimates with ROA(t) as the dependent variable – 2011 to 2020

	Model 1	Model 2	Model 3	Model 4
Variables/Method	Re	Fe	Re	Fe
Constant	9.0849	3.8364	7.4795	2.6285
ESG (t-1)	-0.0002	-0.0169	0.2306***	0.2283***
ESG^2 (t-1)	N.A.	N.A.	-0.0031***	-0.0033***
Log (Total_Assets)	-0.2238	0.6628	-0.3908	0.4011
Debt_Assets	-0.0839***	-0.1052***	-0.0835**	-0.1042**
Cash_Ratio	0.7229***	0.6993***	0.7203***	0.6973**
Sales_Growth	0.0276***	0.0315***	0.0284**	0.0322**
Year (base=2011)				
2012	-0.5289	-0.5508	-0.5268	-0.5419
2013	0.2129	0.1708	0.1644	0.1318
2014	0.2955	0.2490	0.2052	0.1707
2015	0.1806	0.1805	0.0786	0.0925
2016	0.2645	0.2411	0.1240	0.1183
2017	0.3305	0.2558	0.1361	0.0835
2018	1.7309	1.6410***	1.5474**	1.4848**
2019	2.0050	1.8979***	1.8611**	1.7908**
2020	0.7372	0.6594	0.6154	0.5807
GICS Sector (base = Communication Services)				
Consumer Discretionary	3.4589**	Omitted	3.0940**	Omitted
Consumer Staples	4.7124***	Omitted	4.2688***	Omitted
Energy	-3.4921**	Omitted	-3.4436***	Omitted
Financials	1.8614	Omitted	1.8356	Omitted
Health Care	0.8960	Omitted	0.6459	Omitted
Industrials	2.9787**	Omitted	2.6802**	Omitted
Information Technology	3.0374**	Omitted	2.8352**	Omitted
Materials	1.1298	Omitted	0.8467	Omitted
Real Estate	-0.2638	Omitted	-0.4259	Omitted
Utilities	-1.4672	Omitted	-1.7059*	Omitted
R square				
Within	6.60%	6.98%	7.39%	7.7%
Between	16.94%	0.39%	17.11%	1.06%
Total	11.85%	1.9%	12.29%	2.89%
Sigma_u	4.3326	5.7036	4.3428	5.6242
Sigma_e	5.2348	5.2348	5.2149	5.2149
rho	0.4065	0.5427	0.4095	0.5377
Corr(u_i,xb)	0	-0.2531	0	0.0289
Prob > F	0.0000	0.0000	0.0000	0.0000

Note: The dependent variable is ROA(t) for each model.
*** significant at 1% level, ** significant at 5% level, * significant at 10% level.
Re – Random effects, Fe – Fixed effects.
No. of observations – 3380, No. of time periods – 10.
Models 1 through 4 are all executed with robust clustered errors.

References

Ahi, P., Searcy, C., & Jaber, M. Y. (2018). A probabilistic weighting model for setting priorities in assessing sustainability performance. *Sustainable Production and Consumption, 13*, 80-92.

Angrist, J. D., & Pischke, J. (2008). *Mostly harmless econometrics: An empiricist's companion* Princeton university press.

Barnett, M. L., & Salomon, R. M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal, 27*(11), 1101-1122.

Brammer, S., & Pavelin, S. (2008). Factors influencing the quality of corporate environmental disclosure. *Business Strategy and the Environment, 17*(2), 120-136.

Brockett, A., & Rezaee, Z. (2012). *Corporate sustainability: integrating performance and reporting* John Wiley & Sons.

Brogi, M., & Lagasio, V. (2019). Environmental, social, and governance and company profitability: Are financial intermediaries different? *Corporate Social Responsibility & Environmental Management, 26*(3), 576-587. doi:10.1002/csr.1704

Cek, K., & Eyupoglu, S. (2020). Does Environmental, Social and Governance Performance Influence Economic Performance? *Journal of Business Economics & Management, 21*(4), 1165-1184. doi:10.3846/jbem.2020.12725

Cohen, J., Holder-Webb, L., Nath, L., & Wood, D. (2011). Retail investors' perceptions of the decision-usefulness of economic performance, governance, and corporate social responsibility disclosures. *Behavioral Research in Accounting, 23*(1), 109-129.

Dubbink, W., Graafland, J., & Van Liedekerke, L. (2008). CSR, transparency and the role of intermediate organisations. *Journal of Business Ethics, 82*(2), 391-406.

Escrig-Olmedo, E., Muñoz-Torres, M. J., & Fernández-Izquierdo, M. Á. (2013). Sustainable development and the financial system: Society's perceptions about socially responsible investing. *Business Strategy and the Environment, 22*(6), 410-428.

Freeman, R. E., & McVea, J. (2001). A stakeholder approach to strategic management. *The Blackwell Handbook of Strategic Management,* , 189-207.

Friedman, M. (2007). The social responsibility of business is to increase its profits. *Corporate ethics and corporate governance* (pp. 173-178) Springer.

Graafland, J. J. (2002). Sourcing ethics in the textile sector: the case of C&A. *Business Ethics: A European Review, 11*(3), 282-294.

Hart, S. L., & Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment, 5*(1), 30-37.

Hermalin, B. E., & Weisbach, M. S. (2001). Boards of directors as an endogenously determined institution: A survey of the economic literature.

Hill, R. P., Ainscough, T., Shank, T., & Manullang, D. (2007). Corporate social responsibility and socially responsible investing: A global perspective. *Journal of Business Ethics, 70*(2), 165-174.

IEA. (2019a). Energy Efficiency 2019. Retrieved from https://www.iea.org/reports/energy-efficiency-2019

IEA. (2019b). Time to act as global energy efficiency progress drops to slowest rate since start of decade. Retrieved from https://www.iea.org/news/time-to-act-as-global-energy-efficiency-progress-drops-to-slowest-rate-since-start-of-decade

Ionescu, G. H., Firoiu, D., Pirvu, R., & Vilag, R. D. (2019). The Impact of Esg Factors on Market Value of Companies from Travel and Tourism Industry. *Technological & Economic Development of Economy, 25*(5), 820-849. doi:10.3846/tede.2019.10294

King, A., & Lenox, M. (2002). Exploring the locus of profitable pollution reduction. *Management Science, 48*(2), 289-299.

Kurtz, L. (2005). Answers to four questions. *The Journal of Investing, 14*(3), 125-140.

Li, Y., Gong, M., Zhang, X., & Koh, L. (2018). The impact of environmental, social, and governance disclosure on firm value: The role of CEO power. *British Accounting Review, 50*(1), 60-75. doi:10.1016/j.bar.2017.09.007

Lindblom, C. K. (1994). (1994). The implications of organizational legitimacy for corporate social performance and disclosure. Paper presented at the *Critical Perspectives on Accounting Conference, New York, 1994,*

Lokuwaduge, Chitra Sriyani De Silva, & Heenetigala, K. (2017). Integrating Environmental, Social and Governance (ESG) Disclosure for a Sustainable Development: An Australian Study. *Business Strategy & the Environment (John Wiley & Sons, Inc), 26*(4), 438-450. doi:10.1002/bse.1927

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly, 48*(2), 268-305.

McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of Management Review, 26*(1), 117-127.

Mervelskemper, L., & Streit, D. (2017). Enhancing Market Valuation of ESG Performance: Is Integrated Reporting Keeping its Promise? *Business Strategy & the Environment (John Wiley & Sons, Inc), 26*(4), 536-549. doi:10.1002/bse.1935

Minutolo, M. C., Kristjanpoller, W. D., & Stakeley, J. (2019). Exploring environmental, social, and governance disclosure effects on the S&P 500 financial performance. *Business Strategy & the Environment (John Wiley & Sons, Inc), 28*(6), 1083-1095. doi:10.1002/bse.2303

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review, 22*(4), 853-886.

Mittal, V. K., & Sangwan, K. S. (2014). Development of a model of barriers to environmentally conscious manufacturing implementation. *Null, 52*(2), 584-594. doi:10.1080/00207543.2013.838649

Nikolaou, I. E., Tsalis, T. A., & Evangelinos, K. I. (2019). A framework to measure corporate sustainability performance: A strong sustainability-based view of firm. *Sustainable Production and Consumption, 18*, 1-18.

Olley, G. S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*(6), 1263-1297.

Orlitzky, M., & Benjamin, J. D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society, 40*(4), 369-396.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies, 24*(3), 403-441.

Porter, M. E. (1991). America s green strategy. *Reader in Business and the Environment, 33*

Porter, M. E., & Van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. *Journal of Economic Perspectives, 9*(4), 97-118.

Qiu, Y., Shaukat, A., & Tharyan, R. (2016). Environmental and social disclosures: Link with corporate financial performance. *The British Accounting Review, 48*(1), 102-116.

Rajesh, R., & Rajendran, C. (2020). Relating Environmental, Social, and Governance scores and sustainability performances of firms: An empirical analysis. *Business Strategy & the Environment (John Wiley & Sons, Inc), 29*(3), 1247-1267. doi:10.1002/bse.2429

Rezaee, Z., & Tuo, L. (2019). Are the Quantity and Quality of Sustainability Disclosures Associated with the Innate and Discretionary Earnings Quality? *Journal of Business Ethics, 155*(3), 763-786. doi:10.1007/s10551-017-3546-y

Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance, 72*(6), 2505-2550.
Romero, S., Jeffers, A. E., Lin, B., Aquilino, F., & DeGaetano, L. (2018). Using ESG Ratings to Build a Sustainability Investing Strategy. *CPA Journal, 88*(7), 36-43.
Sahut, J., & Pasquini-Descomps, H. (2015). ESG Impact on Market Performance of Firms: International Evidence. *El Impacto De ESG En La Rentabilidad De Mercado De Las Empresas., 19*(2), 40-63. doi:10.7202/1030386ar
Scherer, A. G., & Palazzo, G. (2011). The new political role of business in a globalized world: A review of a new perspective on CSR and its implications for the firm, governance, and democracy. *Journal of Management Studies, 48*(4), 899-931.
Seuring, S. A. (2001). Green Supply Chain Costing Joint Cost Management in the Polyester Linings Supply Chain (pp. 71-80). *GMI, Spring,*
Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production, 16*(15), 1699-1710.
Simaens, A., & Koster, M. (2013). Reporting on sustainable operations by third sector organizations: A signalling approach. *Public Management Review, 15*(7), 1040-1062.
Soytas, M. A., Denizel, M., & Usar, D. D. (2019). Addressing endogeneity in the causal relationship between sustainability and financial performance. *International Journal of Production Economics, 210*, 56-71.
Sternberg, E. (1997). The defects of stakeholder theory. *Corporate Governance: An International Review, 5*(1), 3-10.
Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571-610.
Tamimi, N., & Sebastianelli, R. (2017). Transparency among S&P 500 companies: An analysis of ESG disclosure scores. *Management Decision,*
The Governance and Accountability Institute. (2017). FLASH REPORT: 85% of S&P 500 Index® Companies Publish Sustainability Reports in 2017.
Tilling, M. V. (2004). Some thoughts on legitimacy theory in social and environmental accounting. *Social and Environmental Accountability Journal, 24*(2), 3-7.
Trumpp, C., & Guenther, T. (2015). Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. *Business Strategy and the Environment, 26*(1), 49-68.
Van Duuren, E., Plantinga, A., & Scholtens, B. (2016). ESG integration and the investment management process: Fundamental investing reinvented. *Journal of Business Ethics, 138*(3), 525-533.
Velte, P. (2020). Does CEO power moderate the link between ESG performance and financial performance?: A focus on the German two-tier system. *Management Research Review, 43*(5), 497-520. doi:10.1108/MRR-04-2019-0182
Villena, V. H., & Gioia, D. A. (2020). A more sustainable supply chain companies tend to focus on their top-tier suppliers, but the real risks come lower down. *Harvard Business Review, 98*(2), 84-93.
Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic Management Journal, 18*(4), 303-319.
Wang, Z., & Sarkis, J. (2013). Investigating the relationship of sustainable supply chain management with corporate financial performance. *International Journal of Productivity & Performance Management, 62*(8), 871-888. doi:10.1108/IJPPM-03-2013-0033
Weber, O. (2014). Environmental, Social and Governance Reporting in China. *Business Strategy & the Environment (John Wiley & Sons, Inc), 23*(5), 303-317. doi:10.1002/bse.1785

Xie, J., Nozawa, W., Yagi, M., Fujii, H., & Managi, S. (2019). Do environmental, social, and governance activities improve corporate financial performance? *Business Strategy & the Environment (John Wiley & Sons, Inc), 28*(2), 286-300. doi:10.1002/bse.2224

Yu, E. P., Guo, C. Q., & Luu, B. V. (2018). Environmental, social and governance transparency and firm value. *Business Strategy and the Environment, 27*(7), 987-1004.

DECISION SCIENCES INSTITUTE

Delivering Socially Useful Products and Services to the Disadvantaged: A Resource Dependence Perspective

Feigao Huang
University of Tennessee
fhuang4@vols.utk.edu

Wendy L. Tate
University of Tennessee
Wendy.tate@utk.edu

ABSTRACT

The markets consisting of the world's economically disadvantaged are situated in a drastically different institutional context, creating supply chain uncertainties. Companies need to cope with various uncertainties to help the disadvantaged address basic needs. Interviews with social enterprises and text analysis of inclusive business reports were conducted to understand this phenomenon. Drawing on resource dependence theory, this research uncovered five main sources of supply chain uncertainties and identified six major strategies falling into buffering and bridging types that companies adopted to reduce supply chain uncertainties for social and economic viability while delivering socially useful products and services to the disadvantaged.

KEYWORDS: Economically disadvantaged, Supply chain uncertainty, Resource dependence theory, Buffering, Bridging

DECISION SCIENCES INSTITUTE
The Financial Performance of ISO 14001 Certified Firms:
An Attempt to Explain the Widely-Varying Outcomes in Previous Studies

Beate Klingenberg
FOM Hochschule fuer Oekonomie und Management
Email: beate.klingenberg@fom-net.de

Tom G. Geurts
Bucknell University
Email: tgg004@bucknell.edu

ABSTRACT

There exists a substantial body of literature regarding firms who achieve ISO 14001 certification and its impact on the financial performance of those firms. Since the ISO 14001 framework should reduce waste and energy costs associated with the production process, thus leading to reduced negative environmental impact, one could expect a positive impact on financial performance after ISO 14001 certification. However, there is no consistency in the findings reported in the literature. This paper provides an overview of some of the aforementioned literature, and discusses what could be potential issues explaining the widely varying outcomes that are reported.

KEYWORDS: ISO 14001, financial performance, literature review, methodology, financial metrics

INTRODUCTION

Sustainability is an important topic of discussion in our society. Issues such as global warming, micro plastics, and the disappearance of habitats of essential animals such as bees, are very concerning. Arguably firms are pivotal in this discussion. Firms provide goods and services that society needs and consumes, however in the process of providing them they contribute, or are even the main contributors, to the problems discussed above. Since society needs the products firms provide and they are an integral part of society, it is essential that firms become more sustainable. A possible step to accomplishing environmental sustainability at the firm level can be ISO 14001 certification. The ISO 14001 standard was introduced in 1996 and updated in 2004, and is meant to strengthen a firms' commitment to being (more) sustainable through internal improvement of their environmental efficiency (ISO, 2015). As such, firms that receive the certification have made certain improvements in their production process that makes it more in line with the United Nation's Sustainable Development Goals (United Nations, 2015).
The standard does not create environmental goals or targets, but instead provides guidelines that firms should comply with when it comes to operations that impact the environment. Or differently stated, it attempts to provide a comprehensive framework for the implementation of more sustainable processes into the regular operations of firms through continuous improvement processes (Curkovic & Sroufe, 2011). Indeed, firms have an incentive to pursue ISO 14001 certification if they want to be more sustainable. This will signal to society that they are committed to sustainable business practices that result in less harmful waste, efficient production processed, reduced energy usage (Hart & Ahuja, 1996; Darnall & Kim, 2012), and hence reduced operational cost. Ceteris paribus, less waste and higher efficiency should lead to

improved financial performance (Klassen & McLaughlin, 1996), however improved financial performance can also be achieved through market signaling. Since many in society are concerned about sustainability, firms that can show that they care and go beyond lip-service by achieving ISO 14001 certification should see more demand for their products and/or services compared to their non-certified competitors or should be able to ask a higher price or both (Santi & Marti, 2018). In fact, certified firms signal to outside partners that they comply with the certification requirements, thus reducing information asymmetries and value chain costs (Christmann & Taylor, 2006; Delmas, 2002; Heras Saizarbitoria & Boiral, 2013; Montiel et al., 2012).
It comes therefore as no surprise that at this intersection of operations management and finance, researchers have been looking at whether the implementation of ISO 14001 not only leads to environmental benefits, but also to whether it is financially beneficial for firms (Vries et al., 2012). There is a cost associated with implementing more sustainable processes, related to for example more efficient equipment, which can be more expensive; and to higher labor costs due to closer monitoring, or new equipment that is geared completely towards reducing harmful waste, but in itself does not benefit the production of goods and services (Darnall, 2006; Abisourour, et al., 2020). On the other hand, Wang & Xiao (2016) found that marginal costs of the implementation of ISO 14001 related process changes decrease gradually over time. Clearly, if there is a financial benefit for firms to implement ISO 14001 certification, thus when the financial benefits outweigh the associated costs, firms will be more willing to implement the standard, which would benefit society as a whole. Unfortunately, previous studies that have researched the connection between ISO 14001 and financial performance did not bring about conclusive results, but instead provide widely varying outcomes, as was reported previously, see for example Abisourour et al. (2020) and Treacy et al. (2019). Some studies report "negative" benefits, i.e., the costs outweigh the benefits, to substantial improved financial performance on the other extreme of the spectrum of outcomes. The authors of this paper decided to see why this is the case. They analyzed twenty-six papers and for which a detailed review will be presented in the next section.
The paper continuous as follows: First the authors will report the results from their literature search. In the subsequent section they will discuss in more detail several potential reasons why there are widely varying outcomes. In the final section the authors will suggest the appropriate framework for finding a solution, thus reconciling the widely-varying outcomes, and in turn provide a way forward to definitively answer the question what the impact is of ISO 14001 certification on financial performance.

LITERATURE SEARCH

The literature search was performed in the databases ProQuest, EBSCO, JSTOR and Science Direct. Table 1 summarizes the keywords, showing in Set 1 those used to identify papers that cover ISO 14001 certification, while Set 2 consists of those keywords that aim at identifying research related to financial performance. This resulted in 10 keyword combinations. The search was limited to peer reviewed papers since January 1, 2000, since 1996 marks the introduction of this certification and it took some time before relevant papers were published. This initial search resulted in 119 papers. These were coded with respect to their specific coverage of financial performance and ISO 14001 certification, which brought the number of papers down to 26.

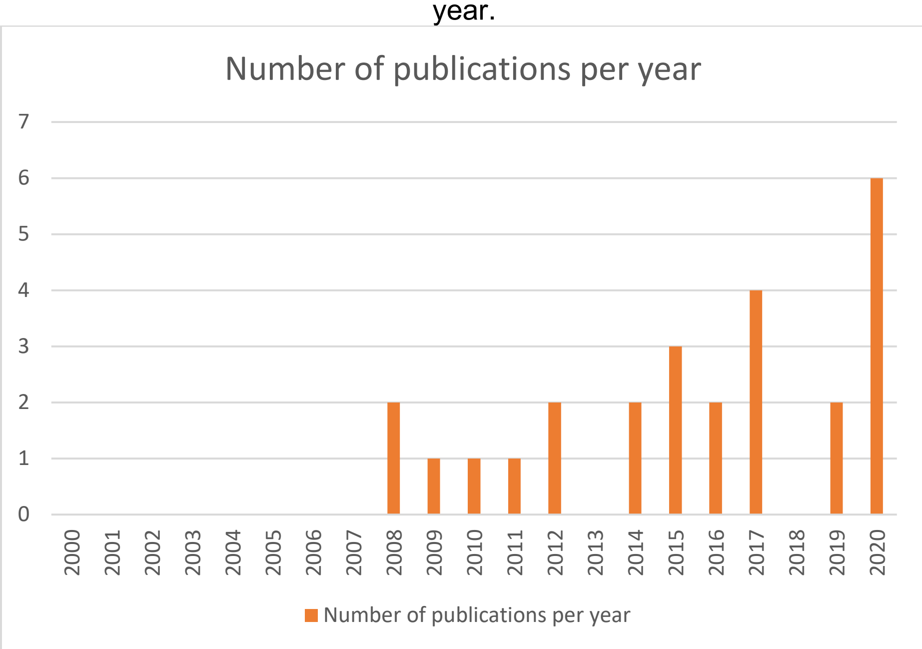
Table 1: Key Words Used for Literature Search

Set 1: ISO 14001	Set 2: Financial Performance
ISO 1400*	Financial performance

ISO 1400*	KPI
	Meas*
	Report
	Performance indicator

Figure 1 show that the first papers appeared in 2008, and that publication rates were relatively constant except an increase in 2020. It is, however, too early to tell whether this indicates an increased (or renewed) interest in the topic.

Figure 1: Publications of the presented search on ISO 14001 and financial performance per year.



Most studies were performed in Asia Pacific, and the Americas (both eight papers). Four studies were performed in Europe, two in the Near East, two in Africa, and one is multi-national. The location of the studies could potentially be of significance for the results, since Marshall & Brown (2003) point out that environmental legislation develops in different institutional environments, where the relationship between public and private sectors can differ greatly. While the ISO 14001 framework is a global standard that is designed to be adapted to geographic, cultural or social conditions (Cary & Roberts, 2011), it is operationalized within the context of national environmental legislation. Wang & Zhao (2020) also point out that there may be differences between developed and developing countries: Firms in developed countries, the creators of the standard, could already pilot the required process changes while the standard was written, and therefore have an advantage over firms in developing countries. This may also create difficulties when comparing research results.
It is important to note that only two separate pairs of the twenty-six papers have the same authorship, which may indicate that the body of literature is (for now) scattered.

ANALYSIS OF PUBLICATIONS
This section discussed four themes that materialized when analyzing the papers, each of them a potential reason for the observed discrepancies in results. They are: 1) How to define financial performance, 2) The methodology used to collect firm data, specifically, how certified firms are

matched with a control firm, and 3) The methodology used to analyze the data, and 4) How to quantify financial performance.

Financial performance
The first issue is that the term “financial performance” is interpreted in different ways, namely on a firm-internal level through the analysis of accounting metrics of financial performance, in general a profitability ratio, or through the analysis of stock market performance. The latter can be seen as a firm-external level analysis, where stock investors judge the impact of certification on financial performance by pushing up the value of the firm in the form of a higher stock price if they judge the certification to be beneficial. If they judge the certification to be unfavorable for the firm they will sell shares, which will result in a decrease of the firms’ market value (Klassen & McLaughlin, 1996; Arts & Vos, 2001; Paulraj & de Jong, 2011).
These two interpretations of “financial performance” may be in itself a reason for the observed incongruence of results: While accounting metrics are designed to provide a direct view of the financial standings of a firm, stock market performance on the other hand is based on the subjective interpretation of said metrics by investors. While the results of quarterly and yearly reports have a strong influence on the judgement of investors, stock markets are also influenced by many other factors. Of course, as was mentioned before, an event such as a certification will have a signaling function, thus leading to a higher valuation of the certified firms’ stock, however, it is the interpretation of the value the firm is creating for the investor, which is not identical to the actual value that the firm is creating. Since the two are not identical, studies that use one measure of financial performance are not automatically comparable to studies that use the other measure of financial performance.

Company matching
A very important methodological procedure that provides the foundation of most studies, and in particular of event-studies, see the next paragraph, is how companies that receive ISO 14001 certification are matched with control groups. In order to perform an analysis that aims at identifying financial performance changes following operational modification, firms that have achieved ISO 14001 certification should be compared with a company that has not yet achieved the certification. However, in order to make a proper comparison this pair of firms should be as similar as possible in all other aspects, hence besides having or not having achieved the certification. Clearly, if there are other industry- or firm-specific differences - besides the certification- then any differentiation in financial performance cannot be unequivocally attributed to presence or absence of certification. Indeed, industry-specific effects should be expected: As Klassen and McLaughlin (1996) point out, “dirtier” industries face higher environmental protection cost and Bouslah et al. (2010) argue that different industries have different environmental issues and face different social pressure regarding how to solve them, it becomes evident that firms should be carefully matched within industries.

According to the seminal paper by Barber & Lyon (1996), company matching to determine abnormal performance should be based on three criteria: Pre-event performance, industry, and firm size. The first criterion, pre-event performance, is important to prevent mean reversion of accounting data (Penman, 1991; Fama & French, 1995) and allows controlling for other factors that may influence firm performance (Barber & Lyon, 1996). The second criterion, industry, is - besides the aforementioned arguments - relevant as the economic status of an industry may account for financial performance changes of up to 20% (McGahan & Porter, 1997). Furthermore, and important in the context of ISO 14001 certification, environmental issues and therefore the requirements and impact of environmental management systems are industry-specific (Klassen & McLaughlin, 1996; Russo & Fouts, 1997; Bouslah et al., 2010). Most papers

use SIC codes as the metric for industry matching. Lastly, firm size may impact operating performance (Fama & French, 1995) and matching by this criterion is important as a critical indicator for stock performance (Fama & French, 1996). Clearly, the results from analyzing improperly matched companies cannot be trusted, since other factors, besides the certification can drive the results, as outlined in this paragraph.
Besides Barber & Lyon (1996), another well-quoted study is by Corbett et al. (2005), which investigates the relationship between ISO 9000 certification and financial performance, and is very detailed in the description and discussion of its methodology. Corbett et al.'s (2005) matching procedure is also derived from Barber & Lyon (1996). Table 2 shows in the first two rows the criteria introduced by these two studies. In the following rows, the other papers that reference Barber & Lyon (1996) and/or Corbett et al. (2005) or entirely different references are listed, together with the actual criteria employed. .

Table 2 Company matching criteria as referenced and as actually employed

References for matching procedure	Matching Variables and Criteria	Author, Year of original publication and of papers that quote either Barber & Lyon, 1996 or Corbett et al., 2005
Barber & Lyon (1996)	Two-digit SIC code, Four-digit SCI code, Two-digit SIC code and similar size (total assets), Two-digit SIC code and Return On Assets Year of match t ROA 90% - 110%	Original procedures by Barber & Lyon, 1996
Corbett et al. (2005)	Two-digit SIC code, total assets, ROA (normalized over industry) Matched by *t*-2 ROA 90% - 110% Assets 50%-200%	Original procedures by Corbett et al., 2005
Barber & Lyon (1996)	Two-digit SIC code Market value of equity Matched by *t*-1 Firm pairs or control firm portfolio Market value of equity 70-130%	Bouslah et al., 2010
Barber & Lyon (1996)	Two-digit SIC code and ROA One-digit SIC code and ROA No industry match and ROA Matched by *t*-2 Firm pairs ROA 90 – 110%	Lo et al., 2012
Barber & Lyon (1996), Corbett et al. (2005)	Two-digit, one-digit or no SIC code match, combined with On Assets, on ROA and assets, on ROA, on change of ROA in year *t*-3 to *t*-2 Matched by *t*-2 Firm pairs or control firm portfolio ROA 90% - 110%	De Jong et al., 2014

	Assets 50%-200%	
Barber & Lyon (1996), Corbett et al. (2005)	Two-digit SIC and ROA One-digit SIC and ROA Only on assets and ROA Matched by *t*-2 Control firm group, ratio 1:2.28 ROA 90 – 110% Asset 30-300%	Treacy et al., 2019
Barber & Lyon (1996)	Two-digit SIC code, One-digit SIC code or no SIC code match and 90 – 110% of performance One Control portfolio Matched by *t* = 0	Ali et al, 2020
Purnanandam & Swaminathan (2004); Fama & French (1997) (for industry sorting)	Certified firms sorted into industry groups, then by sales, then three by Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA) profit margin (3x3 portfolio) and matched with comparable non-certified firms Year of matching not specified Similarity of sales, EBITDA profit	Paulraj & de Jong, 2011
Hendricks & Singhal (2005)	2-digit SIC code One-on-one matching, separately on ROA, assets and ROA and assets (assets normalized over industry) Year for match not specified Close match (no percentage given)	Lee et al., 2014
Collins (2001) and Collins & Porras (2002)	Nature of business Type of products and services Export market ISO9001 adopter Reputation	Nga, 2009

It becomes clear from analyzing the entries of Table 2 that the matching procedures vary from study to study, even when the same paper is referenced in column one as the inspiration for the matching method. This reflects first of all one underlying challenge: Finding matches. As this is not always possible, the requirement of industry match is “softened” by allowing pairs with fewer or no SIC code matches. Some papers report results separated by the strength of industry matching, others aggregate, which weakens the reliability of those results, as industry-specific effects can no longer be filtered out.

Most studies match at year *t-2*, where *t* is the year of the first certification, see the timing of certification discussion in the following paragraph. This is done since the investment to implement process changes required for the certification generally starts in year *t*-1 (de Jong et al., 2014). Another interesting observation is that some papers use control groups (i.e., averaged firm portfolios of non-certified firms), while the procedure proposed by Hendricks & Singhal (2005) suggests one-on-one firm pairs. Lyon et al. (1999) actually argue that such one-on-one matching delivers stronger results.

In conclusion, although the matching procedure is deemed very critical by the authors of all the reviewed papers, there are significant differences, which could account for the widely-varying financial performance outcomes. It is beyond the scope of this paper to judge which would be the "correct" matching procedure. However, it is obvious that – when comparing the findings of studies – one needs to be aware that there are procedural differences that may play a part in explaining the different outcomes.

Methodology applied
Tables 3 to 5 are organized by the methodology used in different papers, and show the metrics that were evaluated, country and industry of study, and key findings. Within a methodological group, the papers are listed in chronological order of publication.

Event studies for abnormal performance
All papers that apply the event-study methodology reference again the seminal paper by Barber & Lyon (1996) who determine how such event studies should be designed, in order to successfully detect abnormal financial performance. Generally speaking, an event study employs accounting-based metrics to evaluate pre- and post-event performance of a firm. The event of study is in this case obviously ISO 14001 certification, which is, unfortunately, not always well defined, since firms may undergo more than one certification within a larger site, or have multiple sites that get certified over time, or just partially. The literature has generally approached this problem by defining the first certification as "the event", as it is expected that effects on firm performance should start to show already after the first certification, not only when all potential sites are certified (Corbett et al., 2005). Two papers are quoted as the original source of this method: Docking & Dowen (1999) and Simmons & White (1999). The former indeed proposes this approach, with the argument that their research aimed at observing market response to certifications over time, hence it made sense to start with the first certification. Simmons & White (1999) actually chose the first date of certification for a site with the best SIC code-match of a product portfolio. Hence, while the approach of choosing the first certification appears to be well-established in the literature (all papers analyzed here us it), there does not seem to be a congruent argument as to why this will yield strong results.
Within the group of papers that use the event-study methodology, most authors chose the Wilcoxon signed-rank test as the statistical test. This also goes back to Barber & Lyon (1996), as this test provides superior results over t-tests. On the other hand, Conover (1999) found that for highly skewed distributions in the data, the sign test is more appropriate. As can be seen, however, from Table 3, only one study employed this test and it is unclear if other papers checked for skewedness or non-normality. The fact that the authors did not explicitly report whether they performed these checks can potentially be one explanation for the widely-varying outcomes in results, since it could be that they used an inappropriate statistical test.
Looking at the results of the various studies, there appears to be some consistency: Financial performance when interpreted as stock market performance is negatively impacted by certification, while performance metrics such as ROA, Return On Sales (ROS), and Sales Over total Assets (SOA) appear to report positive impact of certification. Specifically, both studies on the fashion and textile industries (Lo et al., 2012 and Ali et al., 2020) appear to be consistent in their reporting of positive results. Looking back at Table 2, both papers appear to be using similar matching criteria, except differing in the year chosen for matching and using company pairs or matching against company portfolios.

Table 3: *Event studies*: Statistical test and metrics used, country and industry of study and key findings

Statistical test	Metrics	Country, industry	Key findings	Authors, Year

Skewedness-adjusted t-statistics	Stock market daily returns	Canada, Forest product and paper industries	No short-term effect, long-term effect negative	Bouslah et al., 2010
Wilcoxon signed-rank test	Stock market returns	US, industry specific matching	Abnormal returns significantly negative	Paulraj & de Jong, 2011
Wilcoxon signed-rank test	ROA, ROS, SOA	US, Fashion and Textiles Industry	ROA and ROS improve in 3-year period after certification	Lo et al., 2012
Wilcoxon signed-rank test, sign test, t-test	ROA, ROS, Return on Equity (ROE), asset turnover	US, industry specific matching	Short-term and long-term improvements of ROA	De Jong et al., 2014
Parametric two-tailed t-test	Assets, liabilities, debt-ratio, market-to book ratio	US, industry specific matching	Assets, liabilities, debt-ratio increase in the long-term, but no consistent effect on market-to-book ratio	Lee et al., 2014
Wilcoxon signed-rank test	Fixed assets turnover, employee turnover, ROA, and others	US, industry specific matching	All financial metrics improve, but fixed assets efficiency diminished in the long-term	Treacy et al., 2019
Wilcoxon signed-rank test	ROA, ROS, SOA	Malaysia, Fashion and textile industries	ROA and ROS improve due to cost efficiencies	Ali et al., 2020
Univariant Analysis	ROA	Spain, hospitality, specifically hotels	No difference between certified or non-certified facilities, however certified ones fared better during financial crisis of 2008	Cavero-Rubio et al., 2020

Regressions of accounting metrics

Table 4 summarized past findings that have employed various methods for hypothesis testing or regressions that do not follow the event-study approach. The majority of the studies apply regression methods or models across industries, which may be overly simplistic. The metrics used vary widely, and so do the results. While the event-studies showed some consistency when using ROA and ROS, these findings are not confirmed in the regression models. One interesting argument is raised, namely the impact of combining ISO 14001 certification with other certifications, such as ISO 9001 (Ionaşcu et al., 2017), however, the argument is not further developed in the literature.

Table 4: *Hypothesis testing and regressions*: Statistical test and metrics used, country and industry of study and key findings

Statistical test	Metrics	Country	Key findings	Authors, Year

Chi-square	ROE, inventory turnover/sales, shareholder equity	Malaysia, industry specific matching	Certified firms have higher average ROE, increase of other metrics not supported	Nga, 2009
Switching regression	Indirect measurement of impact of certification through “net benefit” of greenhouse gas emissions, ROA	Japan, not industry specific	Firms with high financial flexibility (no debt overhang) tend to have non-negative net benefits	Hatateka et al., 2012
Panel-data regression	ROE, ROA, ROS, sales, costs, assets, value added, employees	China, industries with different pollution intensities tested separately	Positive impact on sales and costs with similar magnitudes, negligible financial gains	He et al., 2015
Multiple logistics regression	ROE	Taiwan, not industry specific	ISO14001 Certification is the second-best way to improve ROE, after implementing green processes	Luan et al., 2016
Regression model ISO certification as parameter in model to test effect of sustainability reporting on financial performance	Tobin’s Q, Price/Earning (P/E), Price to Book value, ROS, ROE, Return On Capital Employed (ROCE)	India not industry specific	Inconsistent results across metrics	Goel & Misra, 2015
Regression model	ROA	Romania, not industry specific	Certified firms perform better, particularly if they have combination of certifications	Ionaşcu et al., 2017
Endogenous treatment-regression model based on panel data	Indirect measure of profitability using physical capital and labor	Africa, 41 countries	Certification increases productivity	Goedhuys & Mohnen, 2017
Regression model	Earnings per share (EPS)	Indonesia, not industry specific	Certification does not have an effect	Muda & Wahyuni, 2019

Environmental performance and certification as predictor for EPS				

Various other methods
Table 5 summarizes other methods employed by researchers, together with metrics (if applicable), the country and industry of study, and key findings.

Table 5: *Various methods*: Type of method, metrics used, country and industry of study and key findings

Method	Metrics	Country, industry	Key findings	Authors, Year
Perception Survey	Overall financial performance, and other Indicators of Organizational Performance (IOPs) Detailed questions not provided	India, manufacturing, not industry specific, but firm size, export orientation and years of certification considered	Certification does not result in significant changes in IOPs	Padma et al., 2008
Perception Survey	Overall financial Performance, five-point Likert scale to choose between losses and profits	Canada, France, Germany, Hungary, Japan, Norway, USA, manufacturing, not industry specific	Only substantive implementation of standard leads to financial benefits	Ferrón-Vílchez, 2016
Multiple Case Study		Slovenia, six organizations from manufacturing, water management and service/manufacturing industry	Certification leads only to Triple Bottom-line (TBL) improvements if practices are internalized	Maletič et al., 2015
Case Study	Operational costs	Morocco, chemical industry	Costs are tracked with cost-deployment method	Abisourour et al., 2020
Portfolio analysis	Average monthly returns, standard deviation for buy-and-hold strategy, comparison to S&P 500, Domini Social Equity Fund, Winslow Green Growth Investment and iShares KLD 400 Social Index	US, not industry specific	Portfolio of certified firms outperforms comparison stock indices	Sebastianelli et al., 2015

Econometrics model	Profit Margin ROE, EPS, sales growth	Taiwan, manufacturing	All metrics worsen after certification	Lee et al, 2008
Structured equation modeling	Survey, Sales, market share, new market opportunities	Slovenia, industry not specified	Eco-innovation positively effects firm performance	Hojnik & Ruzzier, 2017
Structured equation modeling	Survey, constructs for firm performance follows Seggie et al. (2006), but is not further disclosed	Turkey, manufacturing, only ISO 14001 certified firms as an indication for commitment to environment	Internal environmental orientation has direct positive effect on financial performance; external environmental orientation has indirect positive effect	Zehir & Ozgul, 2020
Semi-structure interviews		United Arab Emirates, private and public organizations, not industry specific	Certification leads to improved organizational efficiencies	Waxin et al., 2020
Difference-in-difference method	Tobin's Q	China, not industry specific	Certification negatively affects financial performance, but the effect decreases with longer certification time, and it can positively impact global competitiveness, firm size and age positively relate to certification-firm performance relationship	Wang & Zhao, 2020

The variety of methods applied to the research question of ISO 14001 and financial performances covers many methods available for social sciences research, and as such provide a rich source of information. Important arguments are developed that are not the focus of the previously discussed papers. For example, two studies find that it is important *how* the standard is implemented: Only symbolically, or substantive, with internalized procedures and approaches. Aspects such as firm size, as brought up by Wang & Zhao (2020) are also considered in event-studies through firm matching (see section above), however, firm age (and with that, operational experience) have generally been ignored in the other papers that were reviewed.
In the end, though, the results are again entirely inconclusive, and as the approaches presented in this section are vastly different, it is not possible to draw any more generalized conclusions.

Financial metrics
The most commonly used metric is ROA, which is generally recognized as a measurement of firm profitability. Barber & Lyon (1996) define ROA as presented in equation 1:

ROA = (operating income)/(average of beginning- and end- period book value of total assets) (1).

Unfortunately, the definition of ROA used by Barber & Lyon (1996) is not universally accepted. More common is the usage of net income in the numerator, where: Net Income = Operating Income - Net Interest Expense - Tax Expense or Benefit. There can be additional subtractions from operating income to get to net income, but those are less common. Intuitively the usage of operating income in the numerator is more logical when the objective is to evaluate the impact of some kind of improvement on operations, because those improvements would not affect interest and tax expenses. Indeed, having a numerator that is “closer” to the actual operating expenses is preferable and therefore definition (1) is a better choice in the context of this research than the alternative, more common, definition. Eight of the reviewed publications use ROA, and of these, four authors refer to definition (1), while two use net income in the numerator, and two did not explain how ROA is defined. None of the studies specify the denominator (assets) clearly, leaving it open if the average of beginning- and end-period values are used, or either one of these. Barber & Lyon point out that often just the end-period values are employed, and dividing operating income just by ending book value of total assets does not appear to change the general trend in results. This potential deviation from the precise definition may therefore not have significant effects on general results.
Lastly, operating income also contains many cash flows that are unaffected by the implementation of ISO 14001, such as administrative costs and depreciation. Hence, even when using operating income, ROA may not be the best ratio to use.
One paper defines ROA differently, see equation 2, following a definition originally developed by Blanchard et al. (1993):

ROA =
(1 - taxrate) x (before-tax operating income - interest receivables and discount premium + interest on securities + depreciation expense) / total assets at the beginning of the year
(2).

Besides the problem of not using the average assets during the year, this definition also includes many cash flows that are unaffected by the implementation of ISO 14001 and is therefore unsuitable to correctly measure its impact on financial performance.

Another commonly used indicator is ROE. However, ROE is inferior to ROA as a financial ratio, because the denominator is the amount of equity that was used to finance the assets, which means that this ratio is <u>also</u> influenced by how a firm is financed. If there is a difference between firms in how the assets are financed- so-called financial leverage- there will be a difference in ROE, which is not due to ISO 14001 certification, but that difference does show up in the comparison. Or differently stated, a difference in ROE between two firms is not just attributable to whether the firm is ISO 14001 certified, but also due to financial leverage. Since financial leverage has a substantial impact on profits, indeed it is often the most important driver of profits, comparing the ROE of firms and drawing conclusions about the impact of ISO 14001 certification appears to be incorrect.

Five papers use ROS, out of which two define it as operating income/sales (these authors also define ROA using operating income in the numerator), and two use net income/sales, again in line with their respective definition of ROA. One paper does not specify ROS. Arguably ROS using operating income as the numerator is a superior ratio, following the arguments above for ROA. Using sales as the denominator is preferable over using (average) assets or (average) equity, because the dollar amount of sales captures the aforementioned potential increased demand for ISO 14001 certified products and/or services, as well as the ability to ask a higher price once certified (Santi & Marti, 2018). It is still not a perfect ratio, because operating income still contains cash flows unaffected by the certification, but it is arguably the best financial ratio of all the ones mentioned in this review.
Two papers use Tobin's Q, a metric for market valuation that was developed by Kaldor (1966) and refined by Tobin & Brainard (1977). It compares market value of existing assets to the replacement cost of said assets. The literature has recognized that this ratio is not easily operationalized, and Chung & Pruitt (1994) suggest approximating it as shown in equation 3:

$$\text{Tobin's Q} = (MVE + PS + DEBT)/TA \quad (3).$$

Where: MVE = market value of equity; PS = liquidation value of the firm's outstanding preferred stock; DEBT = value of firm's stock liabilities net of its short-term assets plus the book value of long-term debt; and TA = book value of total assets of the firm.

This still appears to be a rather complex metric that requires various inputs that may not be readily available from financial reports, which possibly explains why this otherwise very meaningful metric is rarely used. Unfortunately, the two papers that report it do not provide the underlying definition they followed. Furthermore, a conceptual flaw with using Tobin's Q is that it relies on the premise that ISO 14001 certification is reflected in the firm value, which means that the capital providers, thus shareholders, preferred shareholders, and debt providers, correctly value the impact of the certification. This is unlikely, and as discussed in the previous paragraph, studies that look at firm valuation, for example stock market returns, market-to-book ratios or EPS, generally report negative or inconclusive effects of certification.

Indeed, one aspect of all the financial metrics used in the reviewed papers is that they are general profitability or market valuation metrics. The ISO 14001 standard per se does not define any performance objectives (Boiral & Henry, 2012), but rather prescribes practices for efficient environmental management systems with the main objective to reduce the environmental impact of a firm's operations through continuous improvement practices (Curkovic & Sroufe, 2011). Klingenberg et al. (2013, 2021) already pointed out that ROA may not be a very suitable metric to capture the effect of operational activities on firm financial performance. Furthermore, as Treacy et al. (2019) argue, the literature is largely missing the connection to operational performance, and one should question therefore, whether general profitability metrics in fact capture financial effects of ISO 14001 adequately. Consequentially, Treacy et al.'s study also included metrics related to operational efficiencies, such as fixed asset turnover and employee turnover. Abisourour et al. (2020) propose an entirely different approach, namely the use of the cost-deployment framework which is an essential tool of World Class Manufacturing (WCM) (Digalwar & Metri, 2005; Kodali et al., 2004; Nachiappan & Anantraraman, 2006). It brings the quantification of financial impact of certification down to the process level and allows to identify process-specific losses, and therefore enables tracking of cost savings and improvements of overall efficiency and quality (Giovando et al., 2020).
One may therefore carefully conclude that researchers are becoming aware that the metrics used thus far may not be ideal, and that alternatives should be developed and studied. As

argued above, the measurement of financial performance does not appear to be well-defined, and it may be necessary to first develop a financial performance framework that specifically capture the effects of specific events, such as ISO 14001 certification.

SUMMARY OF THE MAIN RESULTS

Based on the literature analysis presented in the previous section, it can be concluded that there are several issues that might explain the wide variety in outcomes of determining the effect of ISO 14001 certification on firm financial performance. Table 6 summarizes the findings.

First of all, the paper discusses that an understanding of "financial performance" differs between using accounting measures or stock market performance as a metric, which are not comparable methods, as the latter looks at the firm from the lense of the investor, while the former analyses the actual financials of the firm.

Secondly, while most of the analyzed papers apply the same method for company matching prescribed by the literature, the operationalization of the criteria varies. Some researchers have been more diligent in finding closely related company pairs, but often the challenge of finding suitable pairs leads to a softening in criteria, particularly for industry match. As pointed out in the previous sections, ISO 14001 implementation may not mean the same in different industries (Klassen & McLaughlin, 1996; Bouslah et al., 2010), so it appears quite likely that studies with poor industry matching may lead to vastly different outcomes than studies that are industry specific, such as the two papers on the fashion and textile industries, which do provide consistent results.
Thirdly, researchers used vastly different methodologies to measure the effect of certification on financial performance. This variety certainly provides a rich body of literature, but also means that results can not necessarily be compared. While event-studies appear to have a strong acceptance as a viable method, regressions, structured equation modeling, econometrics models, case studies, and perception studies (to name the most commonly applied methods) all are recognized social science methodologies. Great care needs to be applied in understanding the underlying assumptions in order to understand if the respective studies truly address how ISO 14001 certification may impact firm financial performance.

Furthermore, statistical testing methods vary. This may be a minor contributor to the fact that results between publications vary greatly, however, it is important to note that adequate statistical methods are paramount to robust testing results.

Finally, the financial metrics employed vary, although there is a preference for using the generic profitability metrics ROA, ROE, and ROS. As discussed in detail in the respective sections, the definitions of these metrics, however, is not consistent across all research. This also raises doubt that the results can be compared. The main weakness, however, in using these metrics is that they do not specifically reflect the impact of operational changes – as the certification would lead to - , but rather capture other firm activities, such as administration, the way the firm is financed, and more. These metrics are therefore not very appropriate for the research question of determining the effect of ISO 14001 certification on firm financial performance.

Table 6: Summary of Key Findings	
Definition of Financial Performance: accounting metrics or stock performance	Different view of the firm's financials
Firm Pairing	Firm performance, size, and industry:

	Procedures established in literature are implemented inconsistently ➔ *industry specific effects may be missed* ➔ *not easily possible to compare results*
Methodology	Vast variety of methods: ➔ *not possible to compare results*
Statistical testing	Some variations in methods: ➔ *rigor of testing not always transparent*
Financial Metrics	Although there is a set of metrics that are consistently used, i.e. ROA, ROE, and ROS, they are inconsistently defined: ➔ *not possible to compare results* ➔ *weakness in capturing operations' specific effects*

LIMITATIONS AND FUTURE RESEARCH: THE COST DEPLOYMENT APPROACH

Clearly, the presented literature analysis has several limitations, foremost the fact that only twenty-six papers were analyzed. It became evident during the literature search that using only ISO 14001 as a main keyword missed a considerable number of papers that more generally speak about Environmental Management Systems (EMSs) or the European Union's Eco-Management and Audit Scheme (EMAS). Some publications may also have been missed because of the cut-off date of January 2000. Lastly, not yet reported on in detail are previous literature reviews that have been performed on ISO 14001, EMS, and financial or environmental performance (see for example: Molina-Azorín et al., 2009; Salim et al., 2018). Neither of those, however, focus on a detailed analysis of the research methodology as is done in this paper. The literature search and analysis are therefore, as of the writing of this paper, ongoing. Nevertheless, the authors belief that the current findings are already worthwhile for bringing them to the attention to the academic community.

Furthermore, and motivated to improve on the reported inconsistencies in the literature, the authors are currently working on an even more in-depth analysis of the financial metrics used in the literature to better understand 1) the exact calculation of metrics used, and 2) suitability of the metrics. As can already be concluded, instead of focusing on those aspects of the production process that are directly affected by ISO 14001 certification, the existing literature predominantly look at generic financial performance. For example, using a metric that employs Net Income measures also other aspects of financial performance, such as changes in how the firm is financed and/or tax treatment and/or rates.

Indeed, new and better financial metrics should look only at changes in waste, energy use, labor costs, maintenance, etc. which are directly affected by the implementation of ISO 14001, see for example Business Benefits (2021). The development of new metrics the authors want to propose is based on using the so-called cost deployment approach, which looks at the specific effect of required process improvements on costs. Or differently stated, the cost deployment framework is a relevant approach to study how processes or production changes, made to render a firm more sustainable, affect financial performance. Giovando et al. (2020) report the results of a case study analyzing the financial impact of the implementation of WCM on a factory. The framework that they used, namely manufacturing cost deployment, is also applicable for analyzing the financial impact of ISO 14001 certification, as was shown by Abisourour et al. (2020). Giovando et al. (2020) define the transformation costs associated with implementing WCM and the authors of this paper will do the same for the implementation of ISO 14001. It then becomes possible to use manufacturing cost deployment, developed first by Yamashina & Kubo (2002), to link these relevant transformation costs directly to the financial statements and derive an appropriate financial ratio. As Yamashina & Kubo (2002) already point out, it is essential to correctly identify the components of manufacturing costs that are relevant and use only those components in the evaluation.

ACKNOWLEDGEMENTS
The authors want to thank Bryan Scutari, a student at Bucknell University, for his diligent assistance during the literature search. This research is made possible through a generous grant from the Institute of Management Accountants.

REFERENCES
Abisourour, J., Mohsine, H., Mounir, B., & Farchi, A. (2020). ISO14001 combined to cost deployment (EMS-CD): A new financial vision, *International Journal of Production Research*, doi: 10.1080/00207543.2020.1790683.

Ali, Q., Salman, A., Parveen, S., & Zaini, Z. (2020). Green behavior and financial performance: Impact on the Malaysian Fashion industry, *SAGE Open, July-September*, doi: 10.1177/2158244020953179.

Arts, F.M., & Vos., E. (2001). The impact of ISO registration on New Zealand firms' performance: a financial perspective, *The TQM magazine, 13*(3), 180-181.

Barber, B.M., & Lyon, J.D. (1996). Detecting abnormal operating performance: the empirical power and specification of test statistics. *Journal of Financial Economics, 41*, 359-399.

Blanchard, O., Rhee, C., Summers, L., (1993). The stock market, profit, and investment. *Quarterly Journal of Economics 108*(1), 115–136.

Boiral, O., & Henri, J.F. (2012). Modelling the impact of ISO 14001 on environmental performance: a comparative approach. *Journal of Environmental Management, 99*(1), 84-97.

Bouslah, K., M'Zali, B., Turcotte, M.-F., & Kooli, M. (2010). The Impact of Forest Certification on Firm Financial Performance in Canada and the U.S. *Journal of Business Ethics, 96*(4), 551–572.

Business Benefits, (2021). IBM uses ISO 14001 to save 6.7% of global energy use. Retrieved from https://business-benefits.org/case-study/ibm-uses-iso-14001-save-6-7-global-energy-use/, May 26, 2021.

Cary, J., & Roberts, A. (2011). The limitations of Environmental Management Systems in Australian Agriculture. *Journal of Environmental Management, 92*(3), 878-885.

Cavero-Rubio, J.A., & Aracelí, A.-M. (2020). Environmental certification and Spanish hotels' performance in the 2008 financial crisis. *Journal of Sustainable Tourism, 28*(5), 771-796.

Christmann, P., & Taylor, G. (2006). Firm self-regulation through international certifiable standards: determinants of symbolic versus substantive implementation. *Journal of International Business Studies, 37*(6), 863-878.

Collins, J. (2001). *Good to Great*. New York: Harper Collins Publishing.

Collins, J., & Porras, J.I. (2002). *Build to last: successful habits of visionary companies*. New York: Harper Collins Publishing.

Conover, W. J. (1999). Practical nonparametric statistics (3rd ed.). New York: Wiley.

Corbett, C.J., Montes-Sancho M.J., & Kirsch, D.A. (2005). The financial impact of ISO 9000 certification in the United States: an empirical analysis. *Management Science, 51*(7), 1046-1059.

Chung K.H., & Pruitt, S.W., (1994). A simple approximation for Tobin's Q. *Financial Management 23*(3), 70-74.

Curkovic, S., & Sroufe, R. (2011). Using ISO 14001 to promote a sustainable supply chain strategy. *Business Strategy and the Environment*, *20*(2), 71-93.

Darnal, N. (2006). Why firms mandate ISO 14001 certification. *Business & Society, 45*, 354-383.

Darnall, N., & Kim, Y. (2012). Which types of environmental management systems are related to greater environmental improvements? *Public Administration Review, 72*(3), 351-365.

Delmas, M. (2002). The diffusion of environmental management standards in Europe and in the United States: an institutional perspective. *Policy Science*, 35, 91-119.

De Jong, P., Paulraj, A., Blome, C. (2014). The financial impact of ISO 14001 certification: top-line, bottom-line, or both? *Journal of Business Ethics, 119*, 131-149.

Digalwar, A.K, & Metri, B.A. (2005). Performance measurement framework for World Class Manufacturing. *International Journal of Applied Management & Technology, 3*(2), 83-101.

Docking, D.S., & Dowen, R.J. (1999). Market interpretation of ISO 9000 registration. *Journal of Financial Research, 12*(2), 147-160.

Fama E.F., French, K. (1995). Size and book-to-market factors in earnings and returns, *Journal of Finance, 50*, 131-155.

Fama E.F., French, K. (1996). Multifactor explanations of asset pricing anomalies, *Journal of Finance, 51*(1), pp. 55-84.

Fama, E.F., & French, K. (1997). Industry cost of equity. *Journal of Financial Economics, 43*(2), 153-193.

Ferrón-Vílchez, V. (2016). Does symbolism benefit environmental and business performance in the adoption of ISO 14001? *Journal of Environmental Management*, *183*(1), 882-894.

Giovando, G., Crovini, C., & Venturini, S. (2020). *Evolutions in manufacturing cost deployment. Global Business and Economics Review*, 22(1/2), 41-52.

Goedhuys, M., & Mohnen, P. Management standard certification and firm productivity: Micro-evidence from Africa. *Journal of African Development, 19*(1), 61-83.

Goel, P., & Misra, R. (2017). Sustainability reporting in India: Exploring sectorial difference and linkages with firm performance. *Vision, 21*(2), 214-224.

Hatateka, T., Kokubu, K., Kajiwara, T., & Nishitani, K. (2012). Factors influencing corporate environmental protection activities for greenhouse gas emission reduction: The relationship between environmental and financial performance. *Environmental Resource Economics, 53*, 455-481.

Hart, S.L., & Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment, 5*(1), 30-37.

He, W., Liu, C., Lu, J., & Cao (2015). Impacts of ISO 14001 adoption on firm performance: Evidence from China. *China Economic Review*, 32, 43.56.

Hendricks, K.B., & Singhal, V. (2005). An empirical analysis of the effect of supply chain disruptions on long run stock price performance and equity risk of the firm. *Productions and Operations Management, 14*(1), 35-52.

Heras Saizarbitoria, I., & Boiral, O. (2013). ISO 9001 and ISO 14001: towards a research agenda on management standards. *International Journal of Management Review 15*(1), 47-65.

Hojnik, J., & Ruzzier, M. (2017). Does it pay to be eco) The mediating role of competitive benefits and the effect of ISO14001. *European Management Journal, 35*(5), 581-594.

Ionaşcu, M., Ionaşcu, I, Săcărin, M., & Minu, M. (2017). Exploring the impact of ISO 9001, ISO 14001 and OHSAS 18001 certification on financial performance: the case of companies listed on the Bucharest Stock Exchange, *Amfiteatru Economic, 19*(44), 166-180.

ISO, (2015). Introduction to ISO1004:2015. Retrieved from https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100371.pdf, May 10, 2021.

Kaldor, N. (1966). Marginal productivity and the macro-economic theories of distribution: comment on Samuelson and Modigliani. *The review of Economic Studies, 33*(4), 309-319.

Klassen, R.D., & McLaughlin, C.P. (1996). The impact of environmental management on firm performance. *Management Science, 42*(8), 1199-1214.

Klingenberg, B., Timberlake, R., Geurts, T.G., & Brown, R.J. (2013). The relationship of operational innovation and financial performance: a critical perspective. *International Journal of Production Economics*, 142(2), 317-323.

Klingenberg, B., Roberts, J., Brown, R.J., Geurts, T.G., Timberlake, R. (2021). Corrigendum to: The Relationship of Operational Innovation and Financial Performance – A critical perspective", *International Journal of Production Economics*, in print.

Kodali, R., Sangwan, K.S. & Sunnapwar, V.K. (2004). Performance value analysis for the justification of World-Class Manufacturing Systems. *Journal of Advanced Manufacturing Systems, 3*(1), 85-102.

Lee, Y.-C., Hu, J.-L., & Ko, J.-F. (2008). The effect of ISO certification on managerial efficiency and financial performance: An empirical study of manufacturing firms. *International Journal of Management*, 25(1), 166-200.

Lee, S.M., Noh, Y., Choi, D., & Rha, J.S. (2014). The effect of ISO 14001 on equity structure. *Industrial Management & Data systems*, 114(6), 979-991.

Lo, C.K.Y, Yeung, A.C.L., & Cheng, T.C.E. (2012). The impact of environmental management systems on financial performance in fashion and textile industries. *International Journal of Production Economics, 135*, 561-567.

Luan, C.-J., Tien, C., & Chen, W.-L. (2016). Which „green" is better? An empirical study of the impact of green activities on firm performance, *Asia Pacific Management Review, 21*(2), 102-110.

Lyon, J., Barber, B. & Tsai, C. (1999). Improved methods for tests of long-run abnormal stock returns. *Journal of Finance, 54*(1), 165-201.

Maletič, M., Podpečan, M., & Maletič, D. (2015). ISO 14001 in a corporate sustainability context: a multiple case study approach. *Management of Environmental Quality, 26*(6), 872-890.

Marshall, R.S., & Brown, D. (2003). Corporate environmental reporting: What's in a metric? *Business Strategy and the Environment, 12*(2), 87-106.

McGahan, A., & Porter, M.E. (1997). How much does industry matter, really? *Strategic Management Journal, July*, 15-60.

Molina-Azorín, J.F., Claver-Cortés, E., López-Gamero, M.D., & Tarí, J.J. (2009. Green management and financial performance: A literature review. *Management Decisions, 47*(7), 1080-1100.

Montiel, I., Husted B.W., & Christmann, P. (2012). Using private management standard certification to reduce information asymmetries in corrupt environments. *Strategic Management Journal, 33*(9), 1103-1113.

Muda, I., & Wahyuni, E. (2019). An analysis of the effect of environmental performance and the implementation of Environmental Management Systems (ISO 14001) on the Issuer Financial Performance, *Calitatea: Access la Success*, 20(168). 113-117.

Nachiappan, R.M, & Anantharaman, N. (2006). Performance measurement of World Class Manufacturing Systems: an overview. *The Icfai Journal of Operations Management*, 21(6), 29-53.

Nga, J.K.H. (2009). The influence of ISO 14000 on firm performance. *Social Responsibility Journal, 5*(3), 1747-1117.

Padma, P., Ganesh, L.S., & Rajendran, C., (2008). A study of the ISO 14000 certification and organizational performance of Indian manufacturing firms. *Benchmarking: An International Journal, 15*(2), 73-100.

Paulraj, A., & de Jong, P. (2011) The effect of ISO 14001 certification announcements on stock performance. *International Journal of Operations & Production Management, 31*(7), 765-788.

Penman, S. (1991). An evaluation of accounting rate of return. Journal of Accounting, Auditing and Finance, 6, 233-255.

Purnanandam, A., & Swaminathan, B. (2004). Are IPOs really overpriced? *Review of Financial Studies, 17*(3), 811-848.

Russo, M.V., & Fouts, P.A. (1997). A resource-based perspective on corporate environmental performance and profitability, The Academy of Management Journal, 40(3), 534-559.

Salim, H.K., Padfield, R., Hansen, S.B., Mohamad, S.E., Yuzir, A., Syayuti, K., Tham, M.H., & Papargyropoulou, E. (2018). Global trends in environmental management system and ISO14001 research. *Journal of Cleaner Production, 170*, 645-653.

Santi, C., & Martí, C. (2018). Implementing service management standards: Motivations and key factors. In: Heras-Saizarbitoria, I., ed., *ISO 9001, ISO 14001, and new management standards*. Cham (Switzerland): Springer, 83-96.

Sebastianelli, R., Tamimi, N., & Iacocca, K. (2013). Improving the quality of environmental management: impact on shareholder value. *International Journal of Quality & Reliability Management*, *32*(1), 53-80.

Seggie, S.H., Kim, D., & Cavusgil, S.T. (2006). Do supply chain IT alignment and supply chain interfirm system integration impact upon brand equity and firm performance? *Journal of Business Research, 59*, 887-895.

Simmons, B.L., & White, M.A. (1999). The relationship between ISO 9000 and business performance: Does registration really matter? *Journal of Managerial Issues, 11*(3), 330-343.

Tobin, J. & Brainard, W.C., (1977). *Asset markets and the cost of capital*. In: B. Balassa & R. Nelson, eds., Economic progress, private values, and public policy: essays in honor of William Fellner, Amsterdam (The Netherlands): North-Holland Publishing, 235-262.

Treacy, R., Humphreys, P., McIvor, R. & Lo, C. (2019). ISO14001 certification and operating performance: A practice-based view. *International Journal of Production Economics*, *208*, 319-328.

United Nations (2015). *Transforming our world: The 2030 agenda for sustainable development, A/RES/70/1*. Retrieved from https://sustainabledevelopment.un.org/post2015/transformingourworld/publication, 25 March, 2019.

Vries, H.J.D., Bayramoglu, D.K., & Wiele, T.V.D. (2012). Business and environmental impact of ISO 14001. *International Journal of Quality and Reliability Management, 29*(4), 425-435.

Wang., Y.-D., & Xiao, X. (2016). Evaluation and prospect of ISO 14001 Environmental management systems standard. *Journal of Social Science Hunan 2,* 155-159.

Wang, J.-X.,& Zhao, M.-Z. (2020). Economic Impacts of ISO 14001 certification in China and the moderating role of firm size and age. *Journal of Cleaner Production, 274*, https://doi.org/10.1016/j.jclepro.2020.123059.

Waxin, M-F., Knuteson, S.K., & Barholomew, A., (2020). Outcomes and key factors of success for ISO 14001 certifications: Evidence from an emerging Arab Gulf country. *Sustainability, 12*(258), doi:10.3390/su12010258.

Yamashina, H., & Kubo, T. (2002). Manufacturing cost deployment. *International Journal of Production Research, 40*(16), 4077-4091.

Zehir, C., & Ozgul, B. (2020). Environmental orientation and firm performance: The mediating mechanism of green innovation. *Research in Business & Social Science, 9*(5), 2147-4478.

DECISION SCIENCES INSTITUTE
Organizational Enablers and Barriers to Environment Protection Activities in Factories

Osam Sato
Tokyo Keizai University
Email: osamsato@tku.ac.jp

Yoshiki Matsui
The Open University of Japan, and Yokohama National University
Email: y.matsui@ouj.ac.jp, ymatsui@ynu.ac.jp

ABSTRACT

Environment protection activities have become critical for factories today. A variety of stakeholders press companies and factories to improve environmental protection. However, we cannot say that all factories tried their best. Organizational enablers and barriers related to the difference among factories. We conducted an empirical study after the literature review. We adopt stakeholder theory and set three hypotheses with organizational quality programs, clearness of the link to requirements, and organizational barriers from the environmental staff's viewpoint. We conducted a set of an empirical study with data in hand. The result supported our three hypotheses.

KEYWORDS: Environment protection, Environmental staff, Factories, Empirical study, Enablers, and Barriers

INTRODUCTION

Environmental Protection

Governments concern environmental protection as a primary policy. Environmental protection is one of three critical parts of sustainability in the industry. EPA (the United States Environmental Protection Agency) requests sustainable manufacturing to manufacturing sectors (EPA, 2020). Ministry of Environment in Japan set a series of regulations and environmental laws (Ministry of Environment, 2021).

Environmental protection is a significant part of Triple Bottom Line (TBL, 3PL: Profits, People, Planet) (Elkington, 1994; Slaper and Hall, 2011; Bergquist et al., 2019), and environmental, social, and governance (ESG) (Amico, 2019; Chams et al., 2021). Parallel pursuit of them is the goal of sustainability. Sustainable Development Goals (SDGs) become a critical standard for world investment financially, and both leading countries and United Nations (UN) set a goal, "2030 Agenda for Sustainable Development" in 2015 (United Nations, 2015). However, "Countries are on track to miss the Sustainable Development Goals (SDGs) relating to environmental protection" (United Nations, 2021).

As governments and societies in industrialized countries concern environmental protection, environmental protection activities become critical to every factory (plant). Mollenkopf et al. (2010) wrote, "Firms are pressured by stakeholders to be more environmentally conscious and to integrate environmental management into their processes and corporate strategies." As a result, environmental protection has become one of the manufacturing objectives for leading factories (Avella et al., 2011). Avella et al. (2011) wrote, "considering environmental protection as an objective for the production department and not just as an external condition imposed on the firm." Bergquist et al. (2019) recognize active environment protection by the industry as "New Road."

However, we cannot say that all factories engage in environment protection activities to a certain extent. The levels of activities differ significantly among factories. Generally, Small and Medium-Sized Enterprises (SMEs) concern environment protection less than large enterprises (Seidel et al., 2008). Villena and Gioia (2020) wrote that "lower-tier suppliers are almost always worse" and "increasing companies' exposure to serious financial, social, and environmental risks." There may be many reasons to explain such differences as top management leadership, the active commitment of employees, clearness of environmental goal, organizational culture, resource limitation such as investment or knowledge, restriction by government, or pressure from customers and NPO and NGO (Seidel et al., 2008; Bergquist et al., 2019; Adamu et al., 2019).

Stakeholders

Sato and Matsui (2020) found impacts of stakeholders, such as top management, employee, customers, and government, to encourage suppliers to improve environmental performance empirically. The study found that the impacts of both governments and customers are indirect through top management as decision-makers, and effects of both top management and employees are direct to encouraging suppliers to improve the environmental performance. Lopez-Gamero, Molina-Azorin, & Clave-Cortes (2009) wrote that managers' assessment of pollution emission affects environment protection activities. Ahn (2015) noted that the number of environmental staff affects the "greening growth of manufacturing firms."

Although stakeholders who impact environmental performance and their relationships became clear, the next step, how they affect environmental performance, is unknown. Among them, motivation of governments, customers, and top managements are rather apparent. Top managements are decision-makers. Governments, customers, and society (including international and local) push top management to environmental management in a plant (Sato and Matsui, 2020).

Research Motivation

However, the impact of employees, i.e., how they affect environmental performance, is not clear. Their motivation includes an order from top management ("boss ordered me"). In addition, an employee is a customer and a person who lives near the factory simultaneously. Because we cannot manage all reasons and factors simultaneously, we focus only on the organizational aspect, i.e., role within a factory, in this study. We classified these reasons as either enablers or barriers.

The process, how employees proceed with environmental management is still unknown. The motivation of this study is to pursue one more step to make it more transparent. Our research goal is to find what employees think and their concerns to proceed with environment management empirically. Because many and a variety of employees are working in a factory, we focus only on environmental (affair) staff who oversee environmental activities in a plant. Focusing on environmental staff is appropriate because of the function. Ahn (2015) found that the number of environmental staff is a positive and significant explanatory factor of environmental performance. Therefore, we study concern of environmental affair staff to engage environment management in a plant.

We focus on a stakeholder, especially, environmental affair staff in a factory, in this study. We employ stakeholder theory (Freeman et al., 2010; Parmar et al., 2010) as our theoretical background. Stakeholder theory insists that "business can be understood as a set of relationships among groups that have a stake in the activities that make up the business" (Parmar et al., 2010). There is much past research that employed stakeholder theory in literature. "From the sustainability perspective, the stakeholder theory plays a key role" (Gao and Xu, 2017). For example, Driessen and Hillebrand (2013) studied stakeholder integration capability in new product development in organizations.

The rest of the paper is structured as follows. The next section introduces our literature review. The following section explains our research hypotheses. Research design, data analysis, data collection procedure, construct operationalization and measurement, hypothesis testing, and results follow. The subsequent section presents a discussion and conclusions. Next, we discuss the implications and contributions of the study findings. Finally, the concluding section highlights some limitations of the study and offers suggestions for future research.

LITERATURE REVIEW

Environmental and Financial Performance

Business factories must pursue profit to survive in the fierce industrial competition in a market. Environmental protection activities, that stakeholders requested, also need investment both financially and human effort. Therefore, the parallel pursuit of both environmental performance (EP) and financial performance (FP) had been a critical research theme. There was much past research about the effect of EP on FP for academics for a long time (Horváthová, 2012).

Some researchers found a negative effect of EP on FP especially when we measure it in emission-base (end-of-pipe) (Wagner, 2005). Other researchers concluded inconclusive influence of firm-level EP on FP and firm competitiveness (Lefebvre, 2003; Ahn, 2015). However, leading scholars and practitioners found that EP is not necessarily in conflict with FP. "Porter (1991) stipulates that better environmental performance may be beneficial for firms since pollution is a sign of economic inefficiency", and the positive relationship "is valid especially in the long-term, as it is likely to take time for firms to re-structure and adjust to new environmental regulations" (Horváthová, 2012). Sustainable innovation needs investment for environmental protection, but it will create and maintain the long-term competitive advantage (Gao and Xu, 2017).

Today, many scholars and practitioners believe that we can pursue both EP and FP simultaneously. The secret to realizing the Porter hypothesis is introducing environmental management systems (EMS) and re-structuring the production system (Horváthová, 2012).

Contribution of Environmental Staff

We apply stakeholder theory and suppose that contribution of environmental staff to the introduction of EMS and realize EP is critical for a factory. Environmental staff is called environmental affair staffs, employee green teams, etc. The activity of employee green teams is "an essential factor for companies aiming to implement and improve environmental management approaches and practices" (Jabbour et al., 2012).

However, we have few studies that explain how environmental staff contribute to it so far as we know. Although critical, environmental staffs hold limited authority and resource for environment management. We believe, they manage to pursue EP with any opportunity and resource available with other department and employees. We suppose three factors that environmental staffs find either opportunity or constraint as:

1. Organizational quality program.
2. Link to requirement
3. Organizational barriers

Organizational Quality Program

Many factories are conducting organizational programs for quality management. Quality department and quality engineers lead it. Maintaining and improving the quality of processes and products are critical practices for factories (Narasimhan & Schoenherr, 2012). It includes statistical quality control, Total Quality Management (TQM) practices/programs (such as quality circles), and lean/six sigma programs.

We suppose there is a positive relationship or compatibility between organizational quality programs and environmental management because both include the cleanness concept. For example, corporate quality programs such as TQM include the 5S (Sort, Set, Shine, Standardize and Sustain) system. Shine in the 5S system pursues workplace cleanness (JICA, 2021). Environment management also seeks the cleanness of the environment.

Mollenkopf et al. (2010) wrote, "lean and green strategies are often seen as compatible initiatives because of their joint focus on waste reduction" (p.15). Also, Mollenkopf et al. (2010) wrote, "firms can implement these strategies concurrently" (p.34). Cham et al. (2021) noted that '"lean" and "green" congregate toward the same targets, as they both incorporate waste reduction techniques and efficiency strategies' (p.2). Wagner (2005) concluded that there is no significant relationship for inputs-based index between EP and FP, but negative relationship for outputs-based index. We find many reports about the compatibility of green and lean (Friedman, 2008). This approach is the green manufacturing practices (Digalwar et al., 2013) or sustainability innovations. "Conceptualizations of sustainability innovations range from perspectives focusing on ecological improvement, labeled as eco-innovation or green product innovation, to viewing sustainability as the integration of all three bottom-line items - social, ecological, and economic - throughout the product life-cycle. There are some other similar concepts" (Juntunen et al., 2018). Therefore, we believe the compatibility of "lean" and "green" (Garza-Reyes, 2015).

Link to Requirements

Factories face many requests from many stakeholders. Governments set regulations for environmental protection. Taxes and regulations pressure factories to act more sustainably (Campos et al., 2017). Committee for Economic Development (2017) wrote, "regulation is a major way in which government influences the U.S. market economy."

Governments publish environmental standards as we wrote in the Introduction in this article. International institutes also publish standards such as International Standards Organization

(ISO) 14001. Companies try to pass them to evade penalties and secure more competitive position (Ahn, 2015).

Here, we need to point out a well-defined measurement scale for a pollutant as a base for formal regulation. Unless effective measurement and scale, environmental staff cannot persuade top management to invest in reducing it. Also, if environmental staff cannot measure how harmful the pollutant is, environmental staff cannot persuade workers in other departments. George Serafeim argued that "metrics and measurements form the basis for internal organizational decision making and investor decisions, and are therefore decisive in incentivizing change" (Bergquist et al., 2019, p.11). Therefore, the measure or scale should be formal and well-accepted worldwide. Usually, it takes time for an institution to define the measure or scale and set a standard because "it is only after a wide variety of actors agree on what is important to monitor and measure that these institutions are able to move on specific policy goals" as noted by Hugh Gorman (Bergquist et al., 2019, p.10).

Customers require a certain level of quality for products. ISO provides some environmental standards such as ISO 14000. Customers employ these standards for supplier selection. Mollenkopf et al. (2010) wrote, "Customers using environmental criteria in supplier selection drive global firms to pursue environmental supply chain strategies to be qualified suppliers for foreign customers." Customers do not pay for environmentally defective and low-quality products. Corporate headquarters and factories usually set environmental policies and practices to coordinate internal processes. They include environmental protection or quality standards. Factory managers and workers must understand them clearly and follow the requests. Following these regulations, demands, and criteria improve environmental protection activities.

Following published and defined environmental protection standards are necessary for factories to protect the environment. Employees need to know what they should do (ARCORO, 2020). However, driving workers to requirements and control their effort is not easy. Even though top management announces requirements, requests, and regulations to workers, workers may not fully understand them or accord them. Indeed, guiding workers to follow the rules is a crucial responsibility for managers. Rademaekers et al. (2012) referred to governmental regulation, lack of access to information and knowledge as external organizational change barriers. Therefore, linking workers to requirements is a prerequisite for environmental protection initiatives.

Organizational Barriers

We suppose that environmental staff may feel some organizational barriers that prevent factories from environmental protection. Cross-functional, top, and middle management buy-in may not be enough, and as a result, the factory may not move to environmental management. Workers may not be motivated enough to invest in environmental protection initiatives. Auden Schendler argued that "Organizational barriers and budget constraints make even simple changes difficult to accomplish," and even though "win-win opportunities exist, but that barriers often block their implementation" (Bergquist et al., 2019, p.9).

Corporate culture, inertia, or resistance to change, weaken environmental protection orientation (Bergquist et al., 2019). Environmental protection initiatives request effort and process change, which are a burden to managers and all workers. Moreover, managers of all levels may not buy in environmental protection and have not integrated it into their business strategy. Auden

Schendler argued that "these difficulties are not caused by foolishness or intransigence," "but by understandable, and possibly unavoidable, conflicting business incentives." (Bergquist et al., 2019) Organizational barriers may be a significant concern for environmental staff if it exists. Rademaekers et al. (2012) referred to corporate culture, lack of top-level commitment, employee participation, and acceptance as internal barriers to organizational change. Therefore, if these organizational barriers are high, the results of environmental protection initiatives may not be high enough.

RESEARCH HYPOTHESES

Based on the above literature review and inference, we picked up three factors, (1) organizational quality programs, (2) link to requirements, and (3) organizational barriers that affect environmental protection initiatives by environmental staff. We set three hypotheses in this study:

H1: The more critical the organizational quality programs, the more a plant engages in environmental protection initiatives.

H2: The clearer the link to requirements for a factory, the more a plant engages in environmental protection initiatives.

H3: The more organizational barriers the environment affairs staff finds, the less a plant engages in environmental protection initiatives.

We can draw these hypotheses into Figure 1. It is our research model.

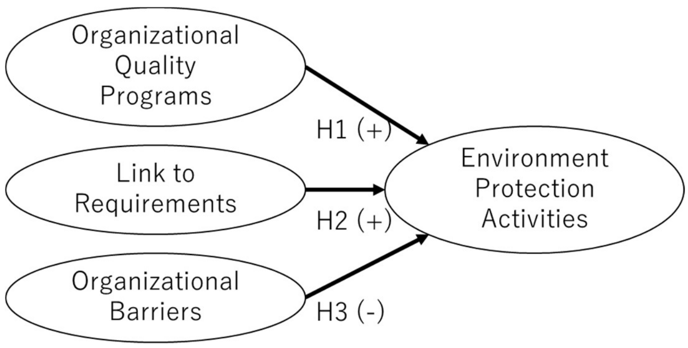


Figure 1: Research Model

RESEARCH DESIGN

Operationalization of Constructs

We employ data from High-Performance Manufacturing (HPM) round 4. Table 1 shows the questions that we used for each construct. Scales for environment protection activities and organizational barriers were a five-point Likert-type scale anchored 1= no extent whatsoever, 3=moderate extent, and 5= very great extent. Questions for organizational programs and environment goals are also five-point Likert scale anchored with 1= completely unimportant, 3=neutral, and 5= extremely important. Environmental affairs staff in a plant responded to these questions.

Table 1: Constructs and Items (Questions)	
Environment Protection Activities: Please indicate the degree to which your plant is engaged in the following initiatives/practices:	
EP01	Water efficiency
EP02	Reducing waste in internal processes (e.g., improving yield or efficiency)
EP03	Improving the workforce environment (e.g., indoor air quality)
EP04	Pollution prevention (eliminating emissions or waste)
EP05	Pollution control (scrubbing, waste treatment)
EP06	Decreasing the likelihood or impact of an environmental accident
EP07	Complying with an industry-wide code of conduct
EP08	Other compliance or auditing program focused on your plant (not on your suppliers)
EP09	Environmental improvements in the disposition of your organization's scrap or excess material (re-use, recycling, etc.)
EP10	Environmental improvements in the disposition of your organization's equipment
Organizational Barriers: To what extent does each of the following limit your plant's ability to engage in environmental initiatives?	
OB01	Insufficient cross-functional buy-in
OB02	Insufficient management buy-in
OB03	Inertia or resistance to change
OB04	Insufficient high level environmental strategy or corporate policy
Organizational Quality Programs: How important is each of the following to your plant's ability to engage in successful environmental initiatives?	
OQ01	Employee green teams
OQ02	Existing TQM practices/programs (such as quality circles)
OQ03	Existing Lean/Six Sigma programs
Link to Requirements: How important is each of the following to your plant's ability to engage in successful environmental initiatives?	
LR01	A clear link between the project and regulatory mandates
LR02	A clear link between the project customer requirements
LR03	A clear link between the project and our own internal policies/goals

DATA ANALYSES AND RESULTS

Data

We employ the fourth-round dataset of the High-Performance Manufacturing (HPM) project (Schroeder & Flynn, 2001; Shimada, T. & Ang, 2014) because it is the latest version available. Each team in our international project visited representative plants and asked/left a standardized questionnaire in each country. Environmental affairs staff in each factory responded to the questionnaire and returned them to our teams. These teams collected the data from factories in electric & electronics, machinery, and transportation equipment. Transportation equipment factory includes automobile assemblers and their parts manufacturers. The fourth round of the HPM project started in 2012-2013, and we collected data in 2014 – 2015 in most countries. Table 1 shows the distribution of sample size by country. The total sample size amounts to 306. However, the actual sample size may be less than that because of the missing values.

Table 2: Data distribution by country					
Country	Count	%	Country	Count	%
Brazil	24	7.8	Japan	22	7.2
China	30	9.8	Korea	26	8.5
Spain	25	8.2	Sweden	9	2.9
Finland	17	5.6	Taiwan	30	9.8
Germany	28	9.2	United Kingdom	13	4.2
Israel	26	8.5	Vietnam	25	8.2
Italy	29	9.5	missing	2	0.7
			Total	306	100

Table 3 shows the sample size by industry.

Table 3: Data distribution by industry		
Industry	Count	%
Electronics	114	37.3
Machinery	110	35.9
Transportation	80	26.1
Missing	2	0.7
Total	306	100.0

Preliminary Data Analysis

First, we conducted exploratory factor analysis (EFA) with all items in Table 1. We conducted EFA to examine construct validity. EFA is a statistical technique that reduces data to a smaller set of factors and explores the underlying theoretical structure of the phenomena without prior specifications of the number of factors and their loadings (Venkatraman, 1989; Statistics Solutions, 2020). We employed a maximum likelihood method with the ProMax rotation in SPSS version 25. We extracted four factors with Eigenvalue criteria, as assumed. The number of data shrinks to 261 because of mission values.

We tested data normality. The Kaiser-Mayer-Olkin (KMO) test of sample adequacy was 0.817, where > 0.9 is marvelous, > 0.8 is meritorious, > 0.7 is middling, and > 0.6 is mediocre (Jacks et al., 2018). Bartlett's Test of Sphericity had an Approximate Chi-Square of 2006.990 at a significance of 0.000. A significance value < 0.05 shows that the data is approximately multivariate normal and suitable for factor analysis.

Table 4 shows the pattern matrix for the four dominant factors, which can be interpreted to be related to environment protection activities, organizational programs, link to requirements, and organizational barriers, respectively.

Table 4: Exploratory Factor Analysis (EFA)				
	Environment Protection Activities	Organizational Programs	Link to Requirements	Organizational Barriers
EP01	0.639	0.001	-0.094	0.119
EP02	0.688	0.017	0.050	0.041
EP03	0.694	0.069	0.052	0.010
EP04	0.845	-0.082	-0.053	-0.092
EP05	0.645	0.050	0.097	0.019
EP06	0.825	0.007	0.030	-0.125
EP07	0.697	0.017	-0.015	0.100
EP08	0.695	0.002	0.051	0.009
EP09	0.739	-0.052	-0.006	-0.159
EP10	0.659	-0.004	-0.125	0.183
OB01	0.054	0.855	0.095	-0.043
OB02	0.027	0.832	-0.001	-0.067
OB03	0.014	0.785	-0.075	0.101
OB04	-0.088	0.765	-0.056	0.005
OQ01	0.002	0.037	0.839	0.015
OQ02	-0.064	0.017	0.841	-0.011
OQ03	0.079	-0.106	0.668	0.068
LR01	0.108	0.060	0.138	0.448
LR02	-0.083	-0.012	0.069	0.848
LR03	0.056	-0.033	-0.080	0.869

Next, we conducted a Confirmatory Factor Analysis (CFA) with Amos 25 (Table 5). CFA is "simultaneously tests the unidimensionality of a set of correlated latent constructs, which helps to establish both unidimensionality and convergent validity" (Wu et al., 2010). We computed Cronbach’s alpha, the average variance extracted (AVE), composite reliability (CR), and the square root of AVE at the bottom of the table to assess convergent and discriminant validity. All AVEs exceeded 0.5 (Chung et al., 2015). All CR and Cronbach’s alpha values exceeded 0.7 (Gefen, Straub, & Boudreau, 2000; Nunnally, 1994). They endorse our constructs' measurement reliability and internal consistency (Cronbach, 1951; Forza & Salvador, 2000).

Table 5: Confirmatory Factor Analysis (CFA)				
	Environment Protection Activities	Organizational Programs	Link to Requirements	Organizational Barriers
EP01	0.607			
EP02	0.707			
EP03	0.682			
EP04	0.774			
EP05	0.662			

EP06	0.773			
EP07	0.738			
EP08	0.724			
EP09	0.688			
EP10	0.685			
OB01		0.810		
OB02		0.887		
OB03		0.555		
OB04			0.742	
OQ01			0.814	
OQ02			0.833	
OQ03				0.824
LR01				0.825
LR02				0.804
LR03				0.793
Cronbach's Alpha	0.909	0.777	0.834	0.884
AVE	0.627	0.701	0.757	0.778
CR	0.944	0.871	0.903	0.933
Sqrt (AVE)	0.792	0.837	0.870	0.882

Table 6 shows correlation coefficients between latent variables and the square root of AVE. The square root of AVE (the value of diagonal elements) should be greater than the inter-correlations of constructs (off-diagonal components) (Chung et al., 2015; Dubey et al., 2019). Table 6 satisfies this condition in all columns. Therefore, we proved discriminant validity (Fornell & Larcker, 1981). The correlation between organizational programs and link to requirements and organizational barriers is negative, as we expected. Instead, it looks strange that the correlation between organizational barriers and organizational programs is positive. Remember that the response came from environmental staff. We suggest that the more a factory proceeds with organizational programs, the more environmental staff found organizational barriers.

Table 6:
correlation coefficients between latent variables and the square root of AVE

	Environment Protection Activities	Organizational Programs	Link to Requirements	Organizational Barriers
Environment Protection Activities	0.792			
Organizational Programs	0.339**	0.837		
Link to Requirements	0.334**	0.388**	0.870	
Organizational Barriers	-0.190**	0.191**	-0.022	0.882

**Correlation is significant at the .01 level (by two-tailed test).

Then, we proceeded to the next step, simultaneous equations modeling (SEM).

Simultaneous Equations Modeling Result

We apply SEM to data with AMOS version 25. Figure 2 shows the result. All standardized coefficients are significant. All coefficients, except H2 (from Link to Requirements to environment protection activities, $p = 0.004$) are strongly significant ($p < 0.001$). Model-fit is good enough. Chi-square = 326.459 (D.F. = 164, $p = 0.000$), CMIN/DF = 1.991, NFI = 0.889, IFI = 0.857, TLI = 0.923, CFI = 0.940, FMIN = 1.070, and RMSEA = 0.057. Chi-square rejects the null hypothesis because of the large sample size (Bagozzi et al., 1991, p.436). “With a large N the statistical power of the Chi-square test may reject good models” (Dion, 2008). “The Chi-Square statistic nearly always rejects the model when large samples are used” (Hooper et al., 2008).

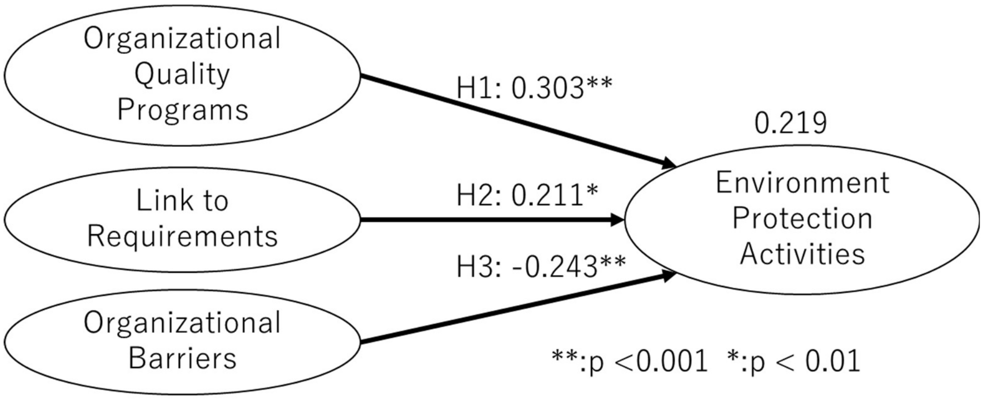


Figure 2: SEM result

DISCUSSION AND CONCLUSIONS

We picked up three explanatory factors base on stakeholder theory to explain environmental protection activities in a factory. Bergquist et al. (2019) describe organizational barriers, missing supportive institutions, the inertia of human systems, and the capacity of the new road as the "Roadblocks to the New Voluntary Road to Environmental Protection" (p.8). Missing supportive institutions relate to lack of "accurate and practical measurement of environmental damage, and the institutions that allow it." It "is a critical determinant of corporate behavior" (p.10). The link to requirement in this study relates to the subject. We include the inertia of human systems as an organizational barrier. The capacity of the new road describes "misalignments between private business incentives and environmental protection" (p.13), which is derived from an economic theory and out of our range of plant-level analysis.

The Conclusions from SEM Result

Today, many stakeholders such as the government, society, customers, and workers request environmental protection activities. Therefore, we adopted stakeholder theory as our base concept. We studied environment protection activities in factories from the environmental staff viewpoint. We set three factors that affect the environment protection activities in factories with data in hand. In this empirical study, because many aspects and factors affect environment protection activities, we focus only on three organizational factors, i.e., organizational quality program, link to requirements, and organizational barriers. Figure 2 shows our result of the SEM model.

We set three hypotheses and conducted an empirical study with SEM.
H1: The more critical the organizational quality programs, the more a plant engages in environmental protection initiatives.
H2: The clearer the link to requirements for a factory, the more a plant engages in environmental protection initiatives.
H3: The more organizational barriers the environment affairs staff finds, the less a plant engages in environmental protection initiatives.

Our result supported all three hypotheses, and we confirmed the following relationships. First, organizational quality programs are the most vital relationship to environmental protection initiatives among two drivers and a barrier. Past literature suggested the relationship, but past studies claim insufficient evidence (Mollenkopf et al., 2010). Second, the effect of the organizational obstacles is negative and substantial. Organizational culture and the buy-in of stakeholders are critical for environmental protection initiatives. Finally, a clear link to requirements is not strong but a statistically significant driver for environmental protection initiatives.

Academic Contribution

Our hypotheses are straightforward, but we do not have enough empirical studies that support these relationships. Our academic contribution is that we add empirical evidence that supports these relationships. Accumulation of empirical evidence is a springboard for a future research stage, and we believe that the publication of this study has a certain level of academic contribution.

Practical Contribution

This study has some practical contributions for managers. First, factories can pursue both organizational quality programs and environmental protection programs simultaneously. The relationship is not a trade-off. Second, making requirements from stakeholders clear to workers is necessary to realize environmental protection initiatives. Stakeholders eager for environmental protection include governments, customers, and employees (Sato and Matsui, 2020). Third, organizational culture and lack of buy-in of managers and employees are significant organizational barriers for environmental protection activities. Managers, especially those in charge of environmental protection initiatives, should remember them.

Research Limitation

There are some research limitations in this study. First, the size of the sample data is not large enough. Our data consists of 261 samples, and it is not enough. We need to repeat this empirical study with more data. Second, we got the data in 2014 - 2015, which is not fresh enough. We need to repeat this study with updated data because the world is changing rapidly. Third, although more internal and external factors affect drivers and barriers for environmental protection activities (Mollenkopf et al., 2010; Rademaekers et al., 2012), we employed only three, i.e., two drivers and one barrier. Fourth, the small coefficient of determination (0.219) suggests that the model's explanatory power is insufficient. We need to add more explaining factors for the larger model. Fifth, although we validated our measurements before our SEM study, we need a more extensive review of our employed measure. Sixth, our model is relatively simple, but we understand that the world is more complex. We need more extensive study both theoretically and empirically. Finally, we need to extend our study to the analysis of detailed mechanisms through which drivers and/or barriers influence each other and on the implementation of environmental initiatives. We could not extend our survey because of time and resource limitations. We need to extend this study and revise the above points, but they are left for future research.

ACKNOWLEDGMENTS

The authors thank the two anonymous reviewers whose comments/suggestions helped improve and clarify this manuscript. This research was sponsored by the JSPS KAKENHI Grant Number 19H01520. The authors thank JSPS for its financial support.

REFERENCES

Adamu, A. A., Wan, C.Y., & Gorondutse, A. H. (2019) Determinants of Sustainable Performance of SMEs: A Proposed Framework, *International Journal of Research and Scientific Innovation*, 6(6), 182-188.

Ahn, P.Q. (2015) Internal Determinants & Effects of Firm-Level Environmental Performance: Empirical Evidences from Vietnam, *Asian Social Science*, 11(4), 190-201.

Amico, A. K. (2019) Investing in the Environment, *Harvard Law School Forum on Corporate Governance*, February 18, https://corpgov.law.harvard.edu/2019/02/18/investing-in-the-environment/.

ARCORO (2020). How to ensure corporate compliance rules are followed by employees, https://arcoro.com/blog/how-to-ensure-corporate-compliance-rules-are-followed-by-employees/

Avella, L., Vazquez-Bustelo, D., & Fernandez, E. (2011). Cumulative manufacturing capabilities: an extended model and new empirical evidence, *International Journal of Production Research*, 49(3) 707–729.

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3), 421-458.

Bergquist, A., Ehrenfeld, J., Schendler, A., Cole, S.A., & King, A. A. (2019) Understanding and Overcoming Roadblocks to Environmental Sustainability. Past Roads and Future Prospects, *Harvard Business School working paper* 19-067.

Campos, J. K., Straube, F., Wutke, S., & Cardoso, P. A. (2017). Creating value by sustainable manufacturing and supply chain management practices – a cross-country comparison, *Proceedings of the 14th Global Conference on Sustainable Manufacturing,* 686-690, doi: 10.1016/j.promfg.2017.02.088

Chams, N., García-Blandón, J. & Hassan, K. (2021) Role Reversal! Financial Performance as an Antecedent of ESG: The Moderating Effect of Total Quality Management, *Sustainability*, 13, 1-18.

Chung, K., Oh, J., Kim, W., & Park, G. (2015). The Effects of Perceived Value of Mobile Phones on User Satisfaction, Brand Trust, and Loyalty. *Advanced Science and Technology Letters*, 114, 10-14.

Committee for Economic Development (2017). Regulation & the Economy: The Relationship & How to Improve It, Retrieved from https://www.ced.org/reports/regulation-and-the-economy, (April 19, 2021).

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests, *Psychometrika* (Vol. 16, pp. 297-334).

Digalwar, A.K., Tagalpallewar, A.R., & Sunnapwar, V.K. (2013). Green manufacturing performance measures: an empirical investigation from Indian manufacturing industries. *Measuring Business Excellence*. 17 (4), 59-75.

Dion, P. (2008). Interpreting Structural Equation Modeling Results: A Reply to Martin and Cullen, *Journal of Business Ethics*, 83(3), 365-368. doi:10.1007/s10551-007-9634-7

Driessen, P. H. & Hillebrand, B. (2013) Integrating Multiple Stakeholder Issues in New Product Development: An Exploration Integrating Multiple Stakeholder, *Journal of Product Innovation Management*, 30(2), 364-379.

Dubey, R., Gunasekaran, A., Childe, S., Papadopoulos, T., & Helo, P. (2019). Supplier relationship management for circular economy. *Management Decision*, 57(4), 767-790. doi:https://doi.org/10.1108/MD-04-2018-0396

Elkington, J. (1994) Towards the Sustainable Corporation: Win-Win-Win Business Strategies for Sustainable Development, *California Management Review*, 36(2), 90-

Environmental Protection Agency. (2020). *Laws & Regulations*. Retrieved from https://www.epa.gov/laws-regulations, (April 19, 2021).

EPA (2020) "Sustainable Manufacturing", Retrieved from https://www.epa.gov/sustainability/sustainable-manufacturing, (April 19, 2021).

Forza, C., & Salvador, F. (2000). Assessing some distinctive dimensions of performance feedback information in high performing plants. *International Journal of Operations & Production Management*, 20(3), 359-385.

Freeman, R. E., Harrison, J. Wicks, A., Parmar, B., & de Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge, UK: Cambridge University Press.

Gao, D. & Xu, Z. (2017). From a systematic literature review to integrated definition for sustainable supply chain innovation (SSCI), *Journal of Cleaner Production*, 142, 1518–1538.

Garza-Reyes, J. A. (2015). Lean and green - a systematic review of the state of the art literature, *Journal of Cleaner Production*,102, 18-29.

Gefen, D., Straub, D., & Boudreau, M. (2000). Structural Equation Modeling Techniques and Regression: Guidelines for Research Practice, *Communications of the Association for Information Systems*, 4, 1-78.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-59.

Horváthová, E. (2012) The impact of environmental performance on firm performance: Short-term costs and long-term benefits? *Ecological Economics*, 84, 91-97.

Jabbour, C. J. C., Santos, F.C.A., Fonseca, S. A., & Nagano, M. S. (2012). Green teams: understanding their roles in the environmental management of companies located in Brazil, *Journal of Cleaner Production*, 46, 58-66.

Jacks, T., Palvia, P., Iyer, L., Sarala, R., & Daynes, S. (2018). An Ideology of IT Occupational Culture: The ASPIRE Values, *The DATA BASE for Advances in Information Systems*, 49(1), 93-117.

JICA (2021), What is 5S principle?, Retrieved from https://www.jica.go.jp/activities/issues/health/5S-KAIZEN-TQM-02/ku57pq00001pi3y4-att/5S_Principle.pdf, (April 19, 2021).

Juntunen, J.K., Halme, M., Korsunova, A. & Rajala, R. (2018). Strategies for Integrating Stakeholders into Sustainability, *Journal of Product Innovation Management*, 36(3), 331-355.

Lefebvre, E., Lefebvre, L., & Talbot, S. (2003). Determinants and impacts of environmental performance in SMEs. *R&D Management*, 33(3), 263-283. http://dx.doi.org/10.1111/1467-9310.00297

Ministry of Environment (2021). Japan's regulations and environmental law, Retrieved from http://www.env.go.jp/en/coop/pollution.html (April 13, 2021).

Mollenkopf, D., Stolze, H., Tate, W. L. & Ueltschy, M. (2010). Green, Lean, and Global Supply Chains, *International Journal of Physical Distribution & Logistics Management*, 40(1/2), 14-41.

Narasimhan, R. & Schoenherr, T. (2012). The effects of integrated supply management practices and environmental management practices on relative competitive quality advantage, *International Journal of Production Research*, 50(4), 1185–1201.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.

Parmar, B., Freeman, R.E., Harrison, J.S. & Colle, S.D. (2010). Stakeholder Theory: The State of the Art, *The Academy of Management Annals*, June, 4(1), 403-445.

Rademaekers, K., Williams, R., Ellis, R., Smith, M., Svatikova, K., & Bilsen, V. (2012). Study on incentives driving improvement of environmental performance of companies, *Report to European Commission - DG Environmen*.

Sato, O. & Matsui, Y. (2020). Stakeholders' Impact on Sustainable Suppliers in Factories, *2020 Annual Conference of the Decision Sciences Institute Proceedings*, pp/416-428, https://decisionsciences.org/wp-content/uploads/2020/11/2020-DSI-Proceedings.pdf

Schendler, A. (2010). *Getting Green Done: Hard Truths from the Front Lines of the Sustainability Revolution*, Public Affairs, New York.

Seidel, M., Seidel, R., Tedford, D., Cross, R. & Wait, L. (2008) A Systems Modeling Approach to Support Environmentally Sustainable Business Development in Manufacturing SMEs, *World Academy of Science, Engineering and Technology*, 48, 121-129.

Schroeder, R. G., & Flynn, B. B. Eds. (2001). *High Performance Manufacturing:* Global *Perspectives*: Wiley.

Shimada, T. & Ang, J. (2014). High Performance Manufacturing (HPM) Project Summary. Retrieved from http://dl.ueb.edu.vn/bitstream/1247/7579/2/Mr%20Chi%20Anh_En.pdf

Slaper, T.F. & Hall, T.J. (2011) The Triple Bottom Line: What Is It and How Does It Work?, *Indiana Business Review*, 86(1), 4-8.

Statistics Solutions. (2020). Exploratory Factor Analysis. Retrieved from https://www.statisticssolutions.com/factor-analysis-sem-exploratory-factor-analysis/

United Nations (2015). Transforming our world: the 2030 Agenda for Sustainable Development, *General Assembly* https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

United Nations (2021). SDGs: Greater urgency needed to meet environmental goals, improved data likely key, *UN News,* https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

Venkatraman, N. (1989). Strategic orientation of business enterprises: The construct, dimensionality, and measurement. *Management Science*, 35(8), 942-962. doi:10.1287/mnsc.35.8.942

Villena, V. H., & Gioia, D. A. (2020). A more sustainable supply chain. *Harvard Business Review*, 98(2), 84-93.

Wagner, M. (2005). How to reconcile environmental and economic performance to improve corporate sustainability: corporate environmental strategies in the European paper industry, *Journal of Environmental Management*, 76(2), 105-118.

Wu, S. J., Melnyk, S. A., & Flynn, B. B. (2010). Operational Capabilities: The Secret Ingredient, *Decision Sciences*, 41(4), 721-754.

DECISION SCIENCES INSTITUTE

What drives consumers to choose their payment method- A study on organized retail outlets?

Shameem Shagirbasha
Great Lakes institute of Management
Email: shawaja@gmail.com

ABSTRACT

The purpose of this research article is to analyse the factors influencing the choice of payment method in organised retail shops. Different methods of payments like cash, debit card, credit card, and mobile wallet payment are available in present scenario, this research article investigated the factors that determine the choice of payment methods in organized retail shops. A mall intercept survey of 357 respondents were collected in retail shops. The respondents were asked to share their retail shop bill details which had method of payment, volume and value of transaction. A structured questionnaire was given to the respondents from which the demographic profile of the respondents were collected. The collected data was then analysed using cluster analysis and decision tree analysis. The results of the analysis revealed the important determinants of the consumers' retail payment mode choice. The value of transaction, income level of the customers and type of retail outlet were identified to be the major determinants of method of payment. The respondents were clustered in to four categories based on their psychological and socio- economic profile. These segments exhibited similar payment behaviour as per the results of cluster analysis.

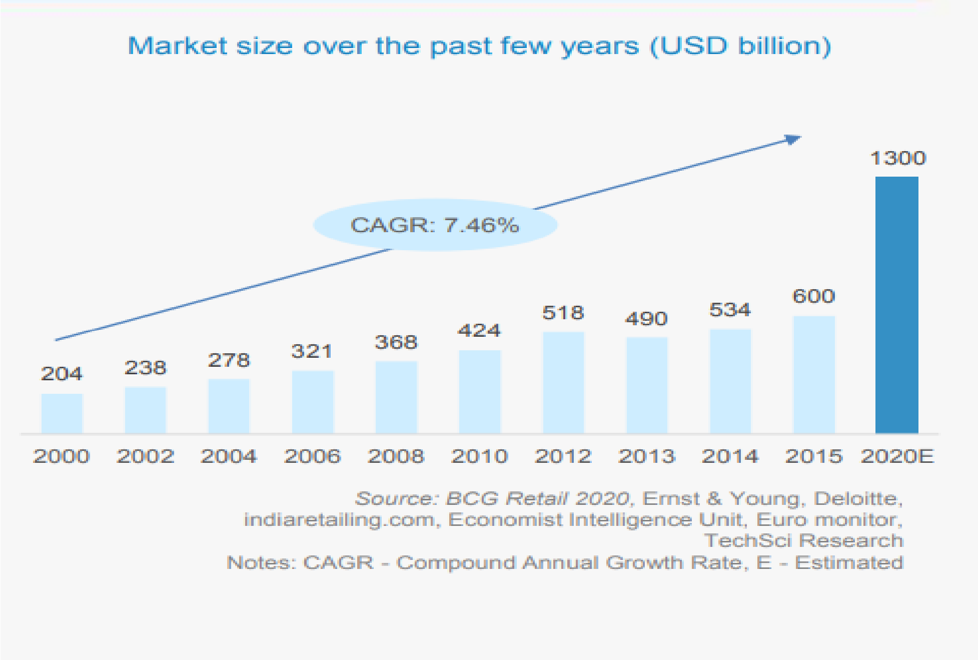
Keywords: payment method, retail shops, consumers, cluster analysis.

INTRODUCTION

World is changing in almost all aspects due to innovation in technology, changes in business and changes in mind-set of the consumers. Retail Markets has seen tremendous change in the way consumers buy products and services. There were cash transactions earlier and that was the predominant way to buy things. The development in technologies has given various payment options to buy like debit cards, credit cards, retail store cards, internet banking, mobile banking, mobile wallets so on and so forth. Now each and every customer has various options in hand to decide to buy products and services. In spite of various options, few customers prefer cash transaction whereas few customers prefer cashless transactions. It is well observed that in economically developed countries and in developing countries, consumers can purchase using more payment options such as cash, debit/credit cards, cheque etc. It actually interests the researchers to understand what makes consumers to select the different payment methods? What are the factors influencing payment mode decisions? Financial institutions, Policy makers, retailers are keen to understand these factors influencing payment mode decisions (van Hove, 2004; King and King, 2005).

The retail sector in India is one of the largest emerging sector in the economy. It is expected that retail industry will grow to USD 1.3 trillion by 2020. The below figure, “Market size over the past few years (USD billion)” will give the growth pattern of the retail sector.

Figure 1: Chart showing market size of organized retail outlets.



Organized retail penetration in India is as low as 8 percent in 2015, expecting it to reach 13 percent by 2019. This indicates high chance of growth for organized retails in India. Among the various categories accounting for the growth, Food & beverage and clothing are the two dominant industries. Hence this study will focus on these two categories to understand the factors influencing the consumers’ payment choice.

LITERATURE REVIEW

There are various factors which influences consumers’ payment mode choice. These include cost characteristics of alternative payment instruments, consumer demographics (King and King, 2005; Amromin et al., 2007), retail transaction size (Jonker, 2007). There are various studies that explains transaction size as key determining factor in retail payment choice (ten Raa and Shestalova, 2004 and Klee, 2008). Handelsman and Munson (1989) found that consumers use cash for low transaction values as compared to high transaction values. Jonker (2007) supported this in his study among Dutch consumers. He found that they are more likely to pay with cash smaller purchase amounts and pay using debit cards for higher purchase amounts. In response to an open-ended question, 8 per cent of their respondents explicitly indicate that they pay cash in certain retail outlets because of the low transaction amounts at these outlets. More recently, Klee (2008) has estimated that a $10 increase in the value of a retail sale decreases the probability of using cash by about 8 per cent.

The recent economic events in India have boosted the efforts towards adopting a less cash economy. According to MasterCard Worldwide Mobile Payments Readiness Index, India averages 33.2. 14% of Indian consumers are familiar with both P2P and m-commerce transactions, and 10% are familiar with POS transactions. However, as data suggests, the usage of cashless payment methods in India has been on a boom. To quote, IMPS usage in retail transactions increased from 78.44m transactions in FY2014-15 to 506.84 in FY2016-17 (source NPCI). This boom is reflected appropriately in other cashless transaction systems as well. This clearly shows that the cashless payment system is really developing in India. However, this usage comes only from the well-organized retail sector. Almost 96% of retail stores in India either are managed by individuals or are family owned. The common practice has been to deal in cash and it is very difficult to change this mindset. Hence, it becomes important to perform an empirical study that measures the impact of adopting cashless payment systems in retail stores. Owing to the fact that retail is a very diverse and dynamic environment, same approach cannot be implemented for all.

Research Question 1: What are the external/environmental factors that drive the choice of payment method in organised retail outlet?

Research Question 2: What are internal factors that drives the choice of payment method in organized retail outlet?

A variety of factors such as consumer demography, income bracket, wallet balance, transaction amount, product category, store location, etc. play an important role in deciding consumer’s payment choice. In this study, we have narrowed down the major factors as transaction size, type of retail and income in deciding payment choice. The results of this study can be used to suggest organized retailers which payment methods to adopt, in order to increase their sales. Based on the literature reviewed the following hypothesis were framed

H1: There is a positive relationship between value of transaction and choice of payment method in organised retail outlets.

H2: There is a positive relationship between type of retail outlet and choice of payment method in organised retail outlets.

H3: There is a positive relationship between income and choice of payment method organised retail outlets.

H4: There is an appropriate match between demographic profile of the retail customers and their choice of payment method in organised retail outlets.

RESEARCH METHODOLOGY

Mall intercept survey of 357 respondents were collected in shopping malls. Prize draw incentives and ten rupees mobile recharge coupons were offered to encourage responses. The respondents were asked to share their retail shop bill details which had method of payment, volume and value of transaction. A structured questionnaire was given to the respondents from which the demographic profile of the respondents were collected. The collected data was then analysed using cluster analysis and decision tree analysis with the help of R and IBM- SPSS software.

The questionnaire was designed using established scales from the previous literatures. The questionnaire had ten different variables which included payment method, value of transaction, Age, Type of retailer, marital status of the respondents, Education, Socio Economic Classification of the respondents, Life style segment of the respondents.

The method of payment was classified in to four categories namely cash payment, credit card payment, debit card payment and mobile wallet payment. The cash payment was a traditional mode of payment and it was included because the literatures suggest that still cash is the king of Indian Economy and no retail shop can rule out cash payment method even though there are sophisticated online and card payment available.

The value of transaction was retrieved from the bill of the sample respondents. The value depicted in the bill was recorded in the questionnaire. The type of the retailer was classified in to two categories namely food and beverages and clothing.

The age of the respondents were divided in to three categories namely young, middle and old age category of people. The young age category of sample respondents were classified as less than 27 years. Sample respondents from 27 years to 45 years were classified as middle age category and the sample respondents of age more than 45 years were classified as old age category.

The income of the sample respondents were classified in to two categories, the sample respondents having less than or equal to Rs. 40,000 as their monthly income and sample respondents who earn more than Rs.40,000 as their monthly income.

The marital status of the sample respondents were classified in to three categories namely unmarried, married and others. The other category of marital status includes widow, widower, divorced, separated etc.,

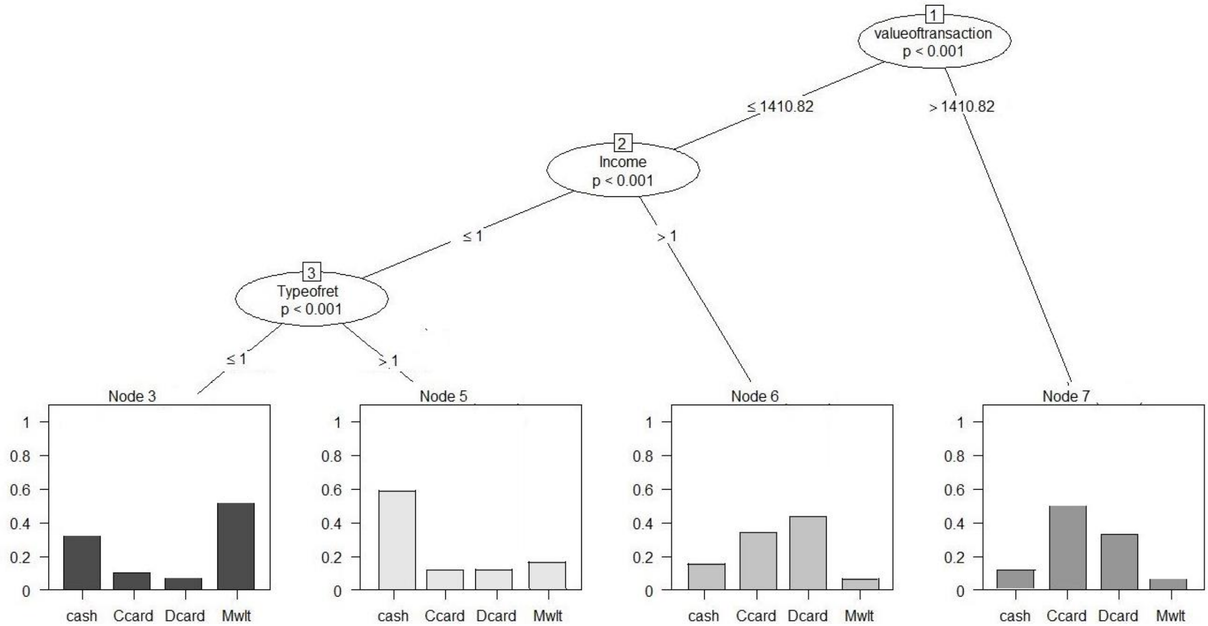
The education qualification was also considered as a demographic variable that determine the choice criteria of the payment process. The education qualification of the sample respondents were classified in to five categories namely high school certificate, diploma, under graduation, post-graduation and professionals. The employment status of the sample respondents were classified in to five categories namely self-employed, employed in public sector, employed in private sector, students and others. The socio economic status of the sample respondents were classified in to three categories namely upper socio economic class, middle socio economic class and lower socio economic class. The lifestyle segment was classified in to three categories based on the activities, interest and opinions of the sample respondents.

Analysis and Interpretation:

Decision tree analysis was carried out to understand the underlying factors that determine choice criteria of payment method. From the decision tree analysis it can be inferred that the value of transaction is the major factor that determine the choice of payment method. The results of decision tree analysis reveal that if the value of transaction is more than Rs. 1410, the sample respondents are opting for card payment. Almost 80 percent of the transactions made over Rs. 1410 were

based on credit cards and debit cards. Thus it can be inferred that the value of transaction plays a major role in deciding the method of payment.
The second node in the decision tree is income classification of the respondents. The income level of the respondents were classified in to two categories. The respondents with more than 40,000 of income and respondents with less than 40,000 income. The results of decision tree analysis revealed that the respondents with more than 40000 income are using credit cards and debit cards frequently when compared to respondents with less than 40,000 income. The respondents with income category less than 40,000 is using cash and mobile wallets more frequently.
The type of retail outlet has a direct influence on the mode of payment. The type of retail outlet was coded as 1 and 2 for the ease of analysis, where 1 represents food and beverages and 2 represents clothing retail. If the value of transaction is less than 1410 and if the retail outlet is food and beverages the sample respondents prefer cash payment.
On the other hand if the value of transaction is less than 1410 and the type of retail outlet is clothing retail outlet the sample respondents prefer mobile wallet payment.

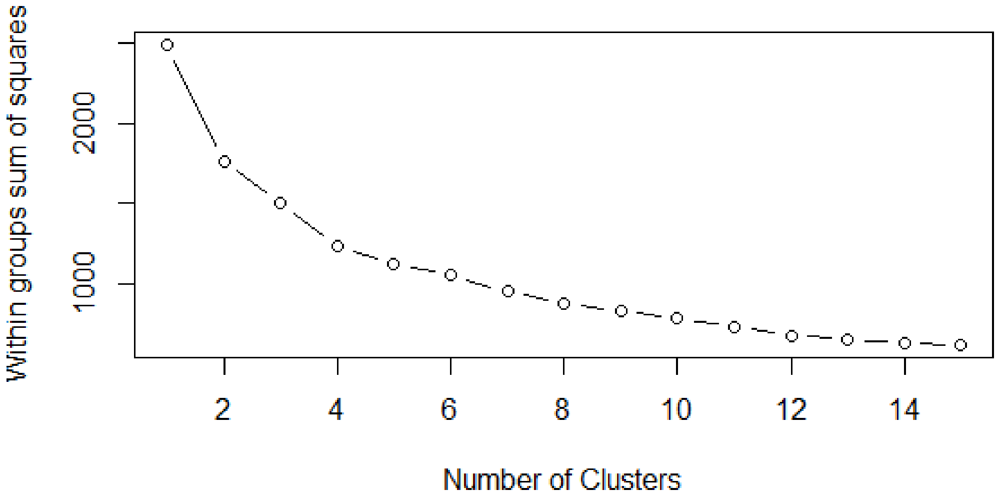
Figure 2: Chart showing Decision tree analysis of the collected primary data



Cluster Analysis

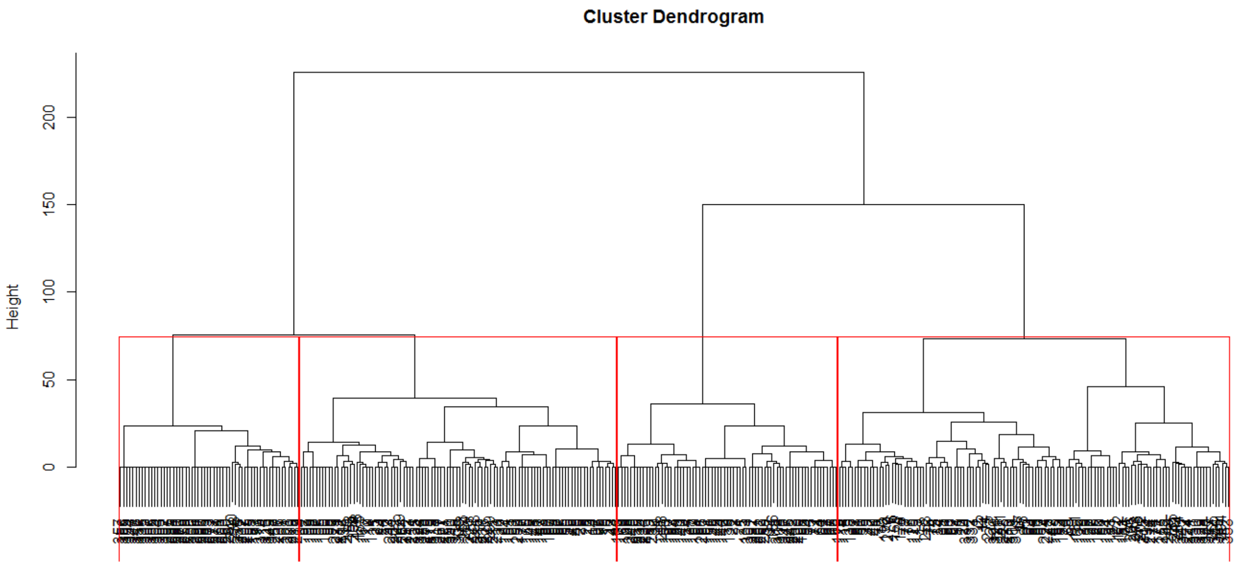
To determine the optimal number of clusters a scree plot was carried out. The results of the cluster analysis scree plot revealed that four clusters can provide optimal results for the model.

Figure 3: Chart showing optimal number of clusters for analysis



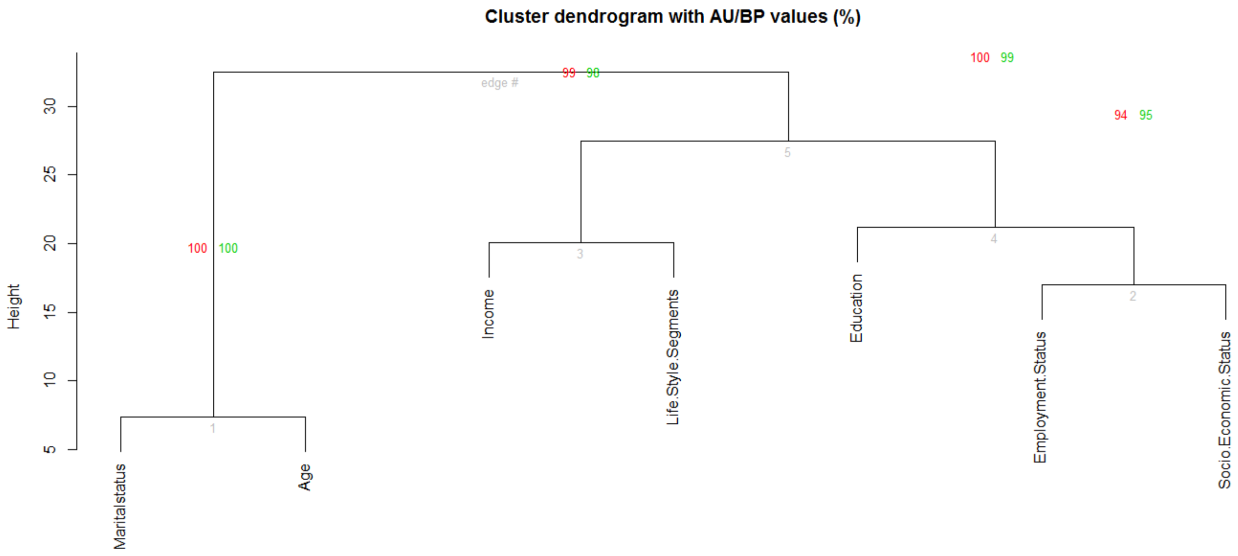
The dendogram results of the cluster analysis was shown in the Figure 4. The results of the dendogram revealed that there are four clusters based on demographic and socio economic classification of the sample respondents.

Figure 4: Dendogram showing four different clusters.



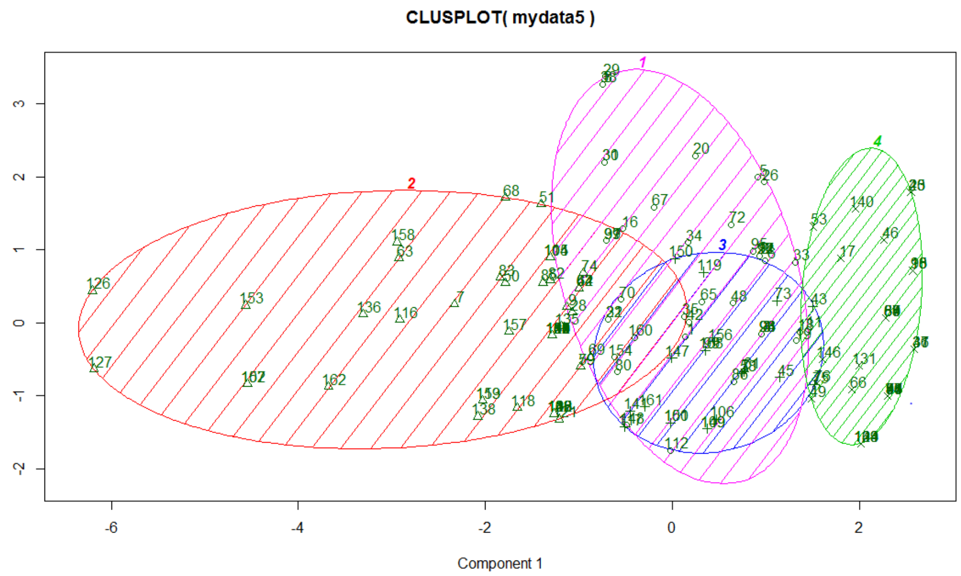
The cluster analysis of the dendogram with AU/BP values represent the variables that show similar characteristics. The age and marital status can be grouped in to a category. The variables income and lifestyle segment can be incorporated in to a category and the education of the sample respondents, employment status of the sample respondents.

Figure 5: Figure showing the variables used for cluster classification.



The cluster plot diagram revealed that the clusters 2 and 4 are distinct, where as there is an overlap between cluster 1 and cluster 3. The clusters were classified based on the lifestyle segment and demographic variables of the sample respondents.

Figure 6: Figure showing the four clusters and the distance between the clusters.



DISCUSSION

From the results of cluster analysis, it could be inferred that there were 4 clusters. The cluster 1 had 99 sample respondents and 27.7 percent of the total sample. The cluster 2 had 62 sample respondents or 17.3 percent of the total sample size, the cluster 3 had 117 or 32 percent of the total sample size. The cluster 4 had 79 sample respondents that corresponds to 22 percent of the total sample respondents. The most choice of payment method for sample respondents in

cluster 1 is credit card and the cluster 1 was found to be dispersed, whereas cluster 3 is more or less like a subset of cluster 1. Thought the number of responses in cluster 3 is comparatively maximum (32.3%), the cluster 3 is found to be homogeneous. Hence it could be inferred that if there is increase in sample size, there could be a clear distinction in cluster 1 and cluster 3.
The sample respondents in cluster 2 are conservative in nature and not so technology savvy. The payment behaviour of such customers has to be observed, hence it would be helpful for the organized retail outlets to device strategies to migrate to cashless payment methods. The sample respondents in cluster 4 are using credit card maximum and they exhibit similar payment behaviour. One sample respondent paid using cash and 52 sample respondents in cluster 2 paid using the credit card. These customers are technologically advanced and hence could use the credit and debit card with ease.

Payment Method	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Cash	40 (11.2%)	48 (13.4%)	46(12.8%)	1 (0.28%)	135 (37.8%)
Credit Card	41(11.5%)	2 (0.5%)	38 (10.6%)	52 (14.5%)	133(37.2%)
Debit Card	9 (2.5%)	4 (1.1%)	12 (3.3%)	19 (5.3%)	44(12.3%)
Mobile Wallet	9 (2.5%)	8 (2.2%)	21(5.8%)	7 (1.9%)	45(12.6%)
Total	99 (27.7%)	62(17.3%)	117 (32.3%)	79 (22.12%)	357 (100%)

MANAGERIAL IMPLICATIONS AND LIMITATIONS

Now-a-days government is encouraging the digital payment methods. E-wallets like Paytm, BHIM are catching the attention of the retailers and most of them has adopted the new payment methods. In this scenario it is important to understand the psychology of the customers and when will the customers choose the different payment methods. Implementing optimal method of payment in retail shops will enhance the business in the retail shops and also boost the spending pattern of the consumers. The research is carried out in organized retail shops and hence it limits generalization of results to other industry.
Conclusion
The research study answered two research questions. The first research question being the internal factors influencing the choice of payment and the second being the external factors that influence the choice of payment. From the study it is clear that the external factors that influence the mode of payment are value of transaction and type of retail outlet and the internal factors that determine the choice of payment method are income, age, life style segment, education etc., In this study, value of transaction emerged as the important factor in influencing the payment choice of consumers. More the income is, more is the usage of payments through cards.

REFERENCES

Arpita Khare, Anshuman Khare, Shveta Singh. 2012. Factors affecting credit card use in India. Asia Pacific Journal of Marketing and Logistics 24:2, 236-256.

Carow, K.A. and Staten, M.E. (1999), “Debit, credit, or cash: survey evidence on gasoline purchases”, Journal of Economics and Business, Vol. 51 No. 5, pp. 409-21.

Chan, R.Y.-K. (1997), “Demographic and attitudinal difference between active and inactive credit cardholders – the case in Hong Kong”, International Journal of Bank Marketing, Vol. 15 No. 4, pp. 117-25.

Chebat, J.C., Laroche, M. and Malette, H. (1988), “A cross-cultural comparison of the attitudes towards and usage of credit cards”, International Journal of Bank Marketing, Vol. 6 No. 4, pp. 42-53.

Danes, S.M. and Hira, T.K. (1990), “Knowledge, beliefs, and practices in the use of credit cards”, Home Economics Research Journal, Vol. 18, pp. 223-35.

Handelsman, M. and Munson, J.M. (1989), “Switching behaviours from credit card to cash payment among ethnically diverse retail customers”, International Journal of Retailing, Vol. 4 No. 1, pp. 31-44.

Hirschman, E.C. (1979), “Differences in consumer purchase behavior by credit card payment system”, Journal of Consumer Research, Vol. 6, pp. 58-66.

Hirschman, E.C. (1982), “Consumer payment systems: the relationship of attribute structure to preference and usage”, Journal of Business, Vol. 55, pp. 531-45.

Humphrey, D.B., Kim, M. and Vale, B. (2001), “Realizing the gains from electronic payments: cost, pricing, and payment choice”, Journal ofMoney, Credit, and Banking, Vol. 33No. 2, pp. 216-34.

Humphrey, D.B., Pulley, L.B. and Vesala, J.M. (1996), “Cash, paper, and electronic payments”, Journal of Money, Credit, and Banking, Vol. 28 No. 4, pp. 914-39.

Jonker, N. (2007), “Payment instruments as perceived by consumers – results from a household survey”, De Economist, Vol. 155 No. 3, pp. 271-303.

Karjaluoto, H. (2002), “Selection criteria for a mode of bill payment: empirical investigation among Finnish bank customers”, International Journal of Retail & Distribution Management, Vol. 30 Nos 6/7, pp. 331-9.

Kaynak, E. and Harcar,T. (2001), “Consumers’ attitudes and intentions towards credit card usage in an advanced developing country”, Journal of Financial Services Marketing,Vol. 6No. 1, p. 24.

Klee, E. (2008), “How people pay: evidence from grocery store data”, Journal of Monetary Economics, Vol. 55 No. 3, pp. 526-41.

Lindley, J., Rudolph, P. and Selby, E. (1989), “Credit card possession and use: changes over time”, Journal of Economic and Business, Vol. 41, pp. 127-42.

Mantel, B. (2000), “Why do consumers pay bills electronically? An empirical analysis”, Economic Perspectives Federal Reserve Bank of Chicago, pp. 32-47 (fourth quarter).

Martell, T. and Fitts, R. (1981), “A quadratic discriminant analysis of bank credit card user characteristics”, Journal of Economics and Business, Vol. 33, pp. 153-9.

Mot, E.S. and Cramer, J.S. (1992), “Mode of payment in household expenditures”, De Economist, Vol. 140 No. 4, pp. 488-500.

Stavins, J. (2001), “Effect of consumer characteristics on the use of payment instruments”, New England Economic Review, pp. 19-31 (third quarter).

DECISION SCIENCES INSTITUTE

Revising Herbert Simon's Decision Making Model

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

ABSTRACT

Herbert A. Simon has made significant contributions in the field decision sciences. While his decision-making model is well grounded theoretically, in this study, we propose revising his decision-making model to integrate Business Analytics (BA) methods and techniques into his model to improve and enhance managers' decision-making effectiveness. We argue that advances in BA and Big Data (BD) have therefore dramatically magnified the need for integrating more recent analytics methods and capabilities into the traditional decision-making models.

KEYWORDS: Herbert Simon, Business Analytics, Decision Making

INTRODUCTION

It may be argued that making rational and sound decisions may be an integral part of any job and a key factor for personal and professional success. However, since business success may depend on the quality of decisions, all relevant information should be taken inro account and a process must be followed when designing a solution to a problem. Similarly, while decision making may involve individuals making personal decisions or a group of experts making complex decisions, one should weigh and examine the positive and negative consequences of their decisions in the context of the problem they are trying to address.

With respect to organizational management, much of what managers do involve making decisions and solving problems. As pointed out by (Simon, 1997), decision-making is the core of management and management is synonymous with decision-making. Delen (2019) makes a similar observation, arguing that managerial decision-making is synonymous with the entire management process. Therefore, managers are constantly called upon to make decisions at every level of management to find the most optimal solution and to make sure organizational goals are achieved. Additionally, making the best decision may be critical for individuals and organizations as the effectiveness and quality of the decisions made by managers may determine failure or success. While decisions made based on relevant information and clearly articulated objectives may bring success, decisions made on the basis of flowed logic and incomplete information may spell failure. Thus, the ability to make decisions in a logical and ordered process is one of the most important primary functions of management.

While managers employ various decision-making models designed to help them make better decisions in a logical and ordered process, none has made more significant contributions as Herbert Simon's decision-making model in the field decision sciences. As argued by Delen (2019), among the related decision-making theories suggested, the one by Simon (1982) stands ahead of the rest and has stood against the time for the past 50 years.

Although Simons’ decision-making model is well grounded theoretically, in this study, we propose revising his decision-making model to incorporate and integrate Business Analytics (BA) and Big Data (BD) capabilities into his model. Given the recent advances that have been made in the BA and BD domains, we argue that his decision-making model should be revised to enhance his decision-making model to allow decision makers to leverage BA methods and tools to improve their decision-making effectiveness, allowing them to make more informed and evidence-based decisions. As argued by (Guszcza and Lucker, 2011) decision makers can improve their traditional decision-making practices through data analytics solutions. Advances in BA and BD have therefore dramatically magnified the need for integrating more recent analytics tools and methods into the traditional decision-making models

LITERATURE REVIEW

Decision-making may be defined as “the process of choosing among two or more alternative courses of action for the purpose of attaining one or more goals” (Delen, 2019). Brousseau et al. (2006) argue that the job of a manager is, above all, to make decisions such as exchanging information, reviewing data, coming up with ideas, evaluating alternatives, implementing directives, and following up.

Decision Making Models

While different models of decision making may lead to dramatically different analyses and predictions (Nitta, 2004), several decision-making models have been proposed in the literature to describe how decisions should be made in a logical and ordered process.

Rational decision-making models consider a number of alternative scenarios and probabilities for each alternative before making decision (Abubakar et al., 2019). During rational decision making, individuals survey alternatives, evaluate consequences from each alternative, and finally do what they believe has the best consequences for themselves (Nita, 2014). Rational decision-making models are based on three assumptions: decision makers have complete and consistent preferences among a set of decision outcomes, they act independently based on full and relevant information, and they always select the decision option that maximizes their utility (Adler, 2020).

The retrospective decision-making models explain how individual rationalize a decision after it has been made (Reed, 2017). Normative models, also known as bounded rationality models, are models in which the chosen alternative is demonstrably the best of all possible alternatives (Delen, 2019). Normative models of decision making include theories such utility, expected-utility theory (EUT), probability, and utilitarianism (Baron, 2004). Intuitive decision-making models argue that people make their decisions by employing their “vague feelings, sense of feeling of pattern or relationships, holistic thinking, immediate insight, seeing the answer without knowing how it was reached” (Abubakar et al., 2019).

Ethical decision-making models describe and integrate the decision-making process and the content variables considered by individuals facing ethical dilemmas (Mc Devitt et al., 2006). Recognition primed decision-making models suggest that people use situation assessment to generate a plausible course of action and use mental simulation to evaluate that course of action and contrast the strengths and weaknesses of recognitional and analytical decision strategies (Klein, 1993).

The ladder of inference decision making model describes the process of belief formation and explains how two people can observe the same raw data yet arrive at very different conclusions (Argyris, 1990). The individual begins with the pool of information available to him/her, then selects some of the information and ignore the rest, interprets the information, drawing on personal or cultural meanings and making assumptions based on those meanings, and finally, draws a conclusion based on that interpretation. Over time, these conclusions inform one's beliefs and drive actions (Levene, 2016). In other words, the individual selectively chooses data to strengthen his or her conviction or belief in the assumption, finally reaching firmly held conclusions (Steen and Sinkai, 2020).

Technically speaking, the paired comparison technique is a research design that yields interval-level scaled scores that are created from ratings made by each respondent for all possible pairs of items under consideration (Lavrakas, 2008). It describes values and compares them to each other. All the potential options are compared, leading to an overview that immediately shows the right decision (Tarrow, 2010). And finally, the phase model of decision-making focuses on successive and distinctive stages in a process, i.e., defining a problem, searching for solutions, and choosing and implementing solutions (Teisman, 2001).

The literature on the decision-making theories and models also makes references to decision making models involving linguistic variables (Delgado et al., 1992), quantitative multicriteria decision models and stochastic decision-making models (Fagundes et al., 2020), computational models of decision making (Busemeyer and Johnson, 2004), single-process models and multiple-strategy models (Sollner et al., 2014), and finally, multi-attribute dynamic decision-making models (Diederich, 1997).

The above decision-making models reflect different fundamental assumptions about human interaction and behavior (Nitta, 2004).

Herbert Simon's Decision-Making Model

Herbert Simon is regarded as a founding father of several of today's most important scientific domains, including the field of decision sciences. His research was noted for its interdisciplinary nature and spanned across the fields of cognitive science, public administration, political science, computer science, and management (Augier and March 2002).

Simon is best known for his theories of bounded rationality and satisficing (Adler, 2020). His rejection of the assumption of perfect rationality led him to develop the concept of bounded rationality (Campitelli and Gobet, 2010), which asserts that people act as perfectly rational actors when making decisions: they are assumed to possess full information about their decision options and consequences, have the ability to define complete and consistent preferences about outcomes, and always chose the option that maximizes utility (Adler, 2020). Simon argues that complex circumstances, limited time, and inadequate mental computational power reduce decision makers to a state of "bounded rationality," (Buchanan and O'Connell, 2006). As a result, people are incapable of choosing an optimal decision. Instead, they are confined to satisfice or select a decision option that appears to be "good enough." (Adler, 2020). In the decision sciences field, Simon's research endeavor aimed to understand the processes that participate in human decision making (Campitelli and Gobet, 2010). He opened a new world of scientific inquiry whose focus is on the development of the most effective and realistic model for the decision makers (Kalantari, 2010). He defined a process model for decision-making, consisting of data gathering, formulating decision options, and choosing a course of action

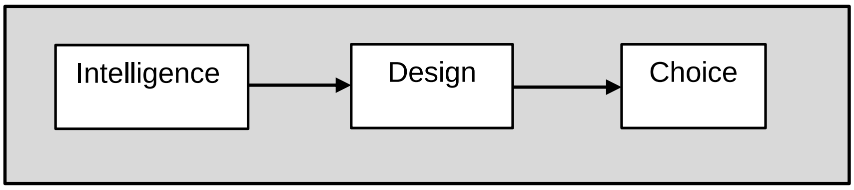
(Adler, 2020). In other words, Simon (1977) argues that a systematic decision-making process involves three major phases: intelligence, design, and choice. While the literature offers a number of different decision-making models, Simon's model is widely accepted as the most concise and yet most complete characterization of rational decision-making (Delen, 2019). As depicted in figure 1, Simon's decision-making model involves three major phases: Intelligence, Design, and Choice.

According to Simon (1977), the decision-making process starts with Intelligence phase, which involves scanning the environment, either intermittently or continuously. It includes several activities aimed at identifying problem situations or opportunities (Delen 2019). In this phase, using their technical and human resources, decision makers collect as much information as possible about the problem they are trying to solve.

The second phase of Simon's decision-making model is the design phase whose main goal is to define and construct a model which represents a system (Markovic, 2018). In other words, the design phase involves designing, developing, and analyzing possible courses of action (Delen, 2019).

Choice is the critical act of decision-making. The choice phase is the one in which the actual decision and the commitment to follow a certain course of action are made. (Delen, 2019). In this phase, a decision is made by selecting and evaluating alternatives defined in the previous step (Markovic, 2018).

Figure 1: Herbert Simon's Decision-Making Model (Simon, 1977).



Integrating Business Analytics Capabilities into Herbert Simon's Decision-Making Model

Several studies conducted in the past attempted to integrate big data analytics into Simon's decision-making model. For instance, Elgendy and Elragal (2016) provided a framework for mapping various big data tools unto Simon's decision-making model. However, their study mainly focuses on specific big data toolsets that may be used by decision makers. Similarly, a study carried out by Poleto et al., (2015) discusses how various technical and organizational elements should be addressed when providing decision makers with decision related opportunities with respect to big data. While it provides a little bit more integrated decision-making model, it nevertheless fails to even discuss how descriptive, predictive, and prescriptive analytics techniques can be employed by decision makers. Finally, Chiheb et al., (2019) offer a theoretical model that attempts to ingrate big data tools into Simon's decision-making model. Although it discusses how big data opportunities can be used to improve and enhance the Intelligence Phase, it fails to elaborate on the last two phases of the Simon's decision-making model.

Although, historically, decision-making has been viewed as a creative and experience-driven, practice, today it is widely perceived as a systematic, evidence-driven, and scientific process.

(Delen, 2019). Therefore, to enhance the likelihood of obtaining the best possible outcomes, the decision-makers are advised to follow a standardized, systematic, and logical decision-making process (Delen, 2019). Employing a proven decision-making model along with a step-by-step approach may be an efficient way to make informed decisions that may have a positive impact on the organization's goals.

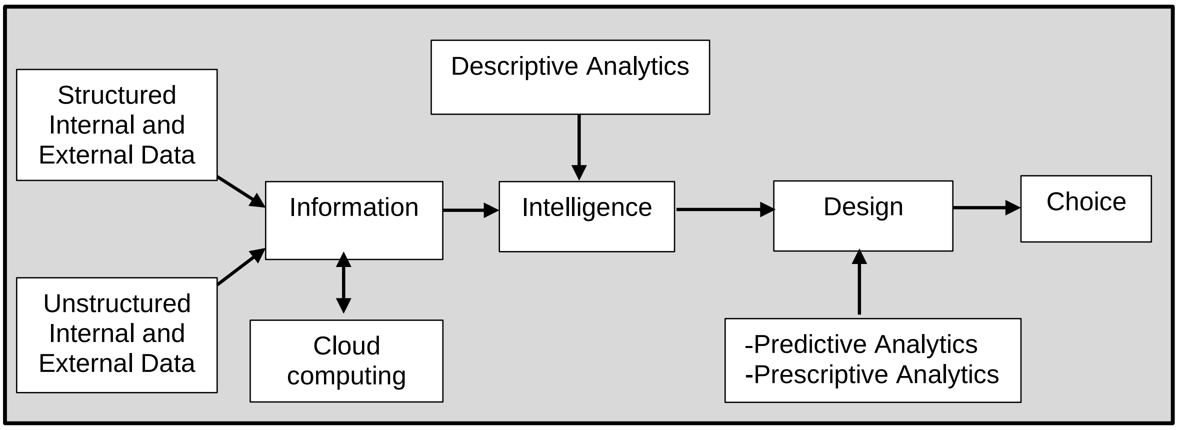
Business Analytics may play an important role in enhancing Simon's decision-making model. As pointed out by (Guszcza, and Lucker, 2011), analytics is the science of better decision-making; and decision-making is the heart of business. As seen in figure 2, organizations collect both structured and unstructured data from a variety of internal and external sources, which may be stored and managed internally using data warehouses. Data collected by the organization may also be stored in a cloud computing environment. As argued by Kotas et al, (2018), recent cloud computing platforms offer features designed for end users with computationally intensive workloads like data analytics.

Structured data refers to the tabular data found in spreadsheets and relational databases (Gandomi and Haider, 2015), which may be generated by humans and machines. Unstructured data on the other hand does not fit into a predefined data model (Ambika, 2020). Unstructured data may be textual or non-textual and human generated are one primary source of big data and are much more challenging to process compared to structure data. Some common examples of unstructured data include text, images, audio, video, and streaming sensor data (Nanjappan et al., 2017). Organizations gathering unstructured data from their environment transform them into structured data by making them readable and comprehensible; and eventually translate them into information by adding meaning and perspective (Figure 2) (Turnbull, 2017).

Simon suggests that people would make rational decisions if only they could gather enough information (Buchanan and O'Connell, 2006). Therefore, decision makers should pay particular attention to the Intelligence Phase as the decision-making process may be dependent upon the right information available to the right people at the right time. As argued by Nita (2014), the keys to a decision are the quality of information about alternatives and individual preferences. Further, management decision making could be adversely affected by unreliable or misleading information (Glantz and Mun, 2011).

In order for managers to design and develop possible courses of actions to solve a problem, they first need to understand the current state of a business or a problem. This can be achieved through descriptive analytics (figure 2). As pointed out by (Delen and Demirkan, 2013), descriptive analytics answer the question of "what happened and/or what is happening?" Hence, in the intelligence phase, managers can employ various descriptive analytics capabilities to understand past and current business performance and make informed decisions (Evans (2012). Simply put, descriptive analytics may allow decision makers to make sense of raw data they collect on internal processes and external business-related activities and helps transform raw data into actionable evidence-based insights in the intelligence phase.

Figure 2: Revised model of Simon's decision-making process



The second phase of Simon's decision-making model involves designing several alternative solutions to the problem identified in the intelligence phase. In this phase, using various predictive and prescriptive analytics applications such as regression analysis, neural network analysis, simulation, and optimization methods, decision makers can identify a set of possible solutions to the problem. As stated by Delen (2019), the alternatives are usually generated through descriptive and predictive levels of analytics by a process of data collection and information creation, and the optimal decision is made through prescriptive analytics. In the design phase, decision makers can further focus their attention on predictive insights hidden in the depths of oceans of data (Guszcza and Lucker, 2011). While predictive analytics uses a variety of models and techniques to predict future outcomes based on historical and current data (Gandomi and Haidar, 2015), prescriptive analytics uses optimization and other techniques to determine which outcome will yield the best result in a given scenario (Gavin, 2019). The role of information is particularly fundamental in the first two phases of Simon's decision making because managers can only choose between alternatives that they know about (Pomerol and Adam, 2004).

Once decision makers gather and generate information about the problem in the intelligence phase through descriptive analytics and generate several possible courses of action in the design phase by employing various predictive and prescriptive analytics methods, they then can make their "choice," which is the last phase of Simon's decision-making model. Integrating descriptive, predictive, and prescriptive analytics capabilities into Simon's decision-making model would eventually help lessen the inefficiencies and cognitive biases that traditional decision-making practices are infused with (Guszcza and Lucker, 2011).

DISCUSSION AND CONCLUSIONS

Herbert Simon argued that a systematic decision-making process involves three major phases: intelligence, design, and choice. While his decision-making model is widely regarded as one of the most important decision-making models in the decision sciences field, in this study, we proposed revising his model to integrate various BA techniques and methods into his model to enhance managers' decision-making effectiveness and aid them in making more accurate and evidence-based decisions.

Making business decisions requires an understanding of the current state of a business. Aided by descriptive analytics capabilities, managers can identify the most pressing issues and problems that their organizations are faced with. To do so, managers would have to translate and transform both structured and unstructured data into readable and comprehensible information, which is the most important ingredient in the design phase.

Simon's model suggests that the key elements of decision making is to identify the best course of action in the design phase. Since, BA involves the use of data to support business decision-making, in the design phase, using various predictive and prescriptive analytics capabilities, managers can identify and generate a set of possible solutions to the problem, and the optimal decision may be identified and implemented through prescriptive analytics.

While no decision-making model is perfect as they all make certain assumptions about human behavior, Simon's model is widely accepted as the most concise characterization of rational decision-making. Having said that, integrating descriptive, predictive, and prescriptive analytics capabilities into Simon's decision-making model would enable decision makers to translate and transform both structured and unstructured data into information at a faster rate, and focus their attention on eliminating cognitive biases by using more accurate information generated in the intelligence phase, eventually allowing managers to make more accurate and informed decisions.

REFERENCES

Abubakar, A., Elrehail, H., Alatailat, M.A., and Elçi, A. (2017). Knowledge management, decision-making style and organizational performance, *Journal of Innovation & Knowledge*, 4, 104-114.

Adler, R. M. (2020). Herbert A. Simon: A biographical sketch, Retrieved May 22, 2021, from https://medium.com/@rich_77042/herbert-a-simon-a-biographical-sketch-959f3d229ce7

Ambika, P. (2020). The digital twin paradigm for smarter systems and environments: The industry use cases*, Advances in Computers*, *117*(1), 321-338.

Argyris, C. (1990). O*vercoming organizational defenses: facilitating organizational learning.* Boston, MA: Allyn and Bacon.

Augier, M., and March, G. J. (2002). A model scholar: Herbert A. Simon (1916–2001), *Journal of Economic Behavior & Organization*, *49*(1), 1-17,

Baron, J. (2004) *Normative models of judgment and decision making,* in Koehler, D.J., and Harvey, N (Eds.), Handbook of Judgment and Decision Making, 19-36, Blackwell Publishing.

Buchanan, L., and O'Connell, A. (2006). A brief history of decision making, Retrieved May 22, 2021, from https://hbr.org/2006/01/a-brief-history-of-decision-making

Brousseau, K.R., Driver, M.J., Hourihan, G., Larsson, R. (2006). The seasoned executive's decision-making style, Retrieved May 23, 2021, from https://hbr.org/2006/02/the-seasoned-executives-decision-making-style#

Busemeyer, J.R and Johnson, J. G. (2004). Computational models of decision making, in Koehler, D.J., and Harvey, N (Eds.), *Handbook of Judgment and Decision Making*, pp.133-154, Blackwell Publishing.

Campitelli G, and Gobet F. (2010). Herbert Simon's decision-making approach: investigation of cognitive processes in experts, *Review of General Psychology*, *14*(4), pp.354-364.

Chiheb, F., Boumahdi, F. and Bouarfa, H. (2019). A new model for integrating big data into phases of decision-making process. *Procedia Computer Science*, 151, 636-642.

Delen, D. (2020). *An overview of the human decision-making process, prescriptive analytics: The final frontier for evidence-based management and optimal decision making,* Pearson FT Press.

Delen, D. and Demirkan, H. (2013). Data, information and analytics as services, *Decision Support Systems, 55*(1), 359-363.

Delgado, M., Verdegay, J. L., and Vila, M.A. (1992). Linguistic decision-making models, *International Journal of Intelligent Systems*, *7*(5), 479-492.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints, *Journal of Mathematical Psychology*, *41*(3), 260-274.

Elgendy, N., and Elragal, A. (2016). Big data analytics in support of the decision-making process, *Procedia Computer Science*, 100, 1071-1084.

Evans, C. (2012). Business analytics: The next frontier for decision sciences, *Decision Line, 43*(2), 4-6.

Fagundes M.V.C, Teles, E.O, Vieira de Melo, SAB, Freires F.G.M, (2020). Decision-making models and support systems for supply chain risk: literature mapping and future research agenda, *European Research on Management and Business Economics*, *26*(2), 63-70.

Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35, 137-144.

Gavin, M. (2019). Business analytics: what it is and why it's important, Retrieved April 29, 2021, from https://online.hbs.edu/blog/post/importance-of-business-analytics.

Glantz, M., and Mun, J. (2011). Building integrated exposure systems, Editor(s): Morton Glantz, Johnathan Mun, 469-487, Credit Engineering for Bankers (Second Edition), Academic Press,

Guszcza, J., and Lucker, J. (2011). Beyond the numbers: Analytics as a strategic capability, Retrieved, May 22, 2021, from https://www2.deloitte.com/uk/en/insights/deloitte-review/issue-8/beyond-the-numbers-analytics-as-a-strategic-capability.html

Kalantari, B. (2010). Herbert A. Simon on making decisions: Enduring insights and bounded rationality, *Journal of Management History, 16*(4), 509-520.

Klein, G. A. (1993). A recognition-primed decision (rpd) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods*, 138-147.

Kotas, C., Naughton, T., and Imam, N. (2018). a comparison of amazon web services and microsoft azure cloud platforms for high performance computing, *IEEE International Conference on Consumer Electronics* (ICCE), 1-4. Las Vegas, NV.

Lavrakas, F. (2008). Paired comparison technique, encyclopedia of survey research methods, *SAGE Publication*.

Levene, G. (2016). Solving the problem with problem-solving meetings, Retrieved May 23, 2021, from https://professional.dce.harvard.edu/blog/solving-the-problem-with-problem-solving-meetings/

McDevitt, R., Giapponi, C. & Tromley, C. (2007). A model of ethical decision making: The integration of process and content, *Journal of Business Ethics*, 73, 219-229.

Nanjappan, V., Liang, H.N., Wang, W., and Man, KL. (2017). Big data: A classification of acquisition and generation methods, in big data analytics for sensor-network collected intelligence, Edited by Chuan-Yu Chang, Hui-Huang Hsu, and Ching-Hsien Hsu, Academic Press; 1st edition.

Nitta, K. (2014). Decision making. *Encyclopedia Britannica*, 24 Jan. 2014, Retrieved May 23, 2021, from https://www.britannica.com/topic/decision-making.

Poleto, T., Carvalho, V.D., and Costa, A. (2015). The roles of big data in the decision-support process: an empirical investigation, In: Delibasic B. et al. (eds) *Decision Support Systems V – Big Data Analytics for Decision Making*. ICDSST 2015. Lecture Notes in Business Information Processing, 216. Springer, Cham.

Pomerol, J., and Adam, F. (2004). Practical decision making – from the legacy of Herbert Simon to decision support systems, decision support in an uncertain and complex world: The IFIP TC8/WG8.3 International Conference.

Reed, G. (2017). How to use decision theory to out-present your competitors, Retrieved May 23, 2021, from https://ethos3.com/2017/04/how-to-use-decision-theory-to-out-present-your-competitors/

Simon, H.A., 1977, *The new science of management decision*, Revised edition; Prentice-Hall, Englewood Cliffs, NJ.

Simon, H. (1997). *Administrative behavior: A study of decision-making behavior in administrative organizations* (4th ed), The Free Press, New York, NY.

Sollner, A., Bröder, A., Glockner, A., Betsch, T., (2014). Single-process versus multiple-strategy models of decision making: Evidence from an information intrusion paradigm, *Acta Psychologica*, 146, 84-96.

Steen. A, and Shinkai K. (2020). Understanding individual and gender differences in conflict resolution: a critical leadership skill, *International Journal of Women's Dermatology, 6(*1), 50-53.

Tarrow, S. (2010). The strategy of paired comparison: toward a theory of practice, *Comparative Political Studies*, 43(2), 230-259.

Teisman, G. R. (2002). Models for research into decision-making processes: on phases, streams and decision-making rounds, *Public Administration*, *78*(4), 937-956.

Turnbull, S. (2017). Grounding a theory of firms in the natural sciences. *Humanistic Management Journal, 1*(2), 159–186.

Croston's Method: 50 years of forecasting intermittent demand

DECISION SCIENCES INSTITUTE

Croston's Method: 50 years of forecasting intermittent demand

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

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

ABSTRACT

In 1972 J.D. Croston proposed an elegant solution to overcoming limitations in forecasting demand using exponential smoothing. Since that publication almost 1000 articles have cited his paper "Forecasting and Stock Control for Intermittent Demands." Furthermore, it has become a standard methodology for forecast models used in inventory stock control systems when intermittent demand is present. This paper analyzes the 947 papers from Google Scholar that cited the work since it was published. An analysis using text mining methodology provides insight into the types of publications that have researched Croston's technique and identify research clusters.

KEYWORDS: Croston, Single Exponential Smoothing, Contextual Analysis

INTRODUCTION

Croston's method (Croston, 1972, as corrected by Rao, 1973) is a forecasting procedure which creates two time series, one for demand and one for the interval between demands. It has become the dominant forecasting method for intermittent demand (Syntetos & Boylan, 2005) since its introduction in 1972. Motivated by supply chain management problems related to products having a high proportion of zero values in their demand history, publications using Croston's method address the problems of inherent demand and its challenge in creating forecasting models. In the 50 years since its publication Croston's article has been cited by at least 934 papers according to Google Scholar. Our paper will analyze the abstracts of the papers that have cited Croston (1972) using text mining methodology to determine the impact this seminal paper has had on methodology involving forecasting intermittent demand.

LITERATURE REVIEW

While no forecasting method has been shown to be superior for all cases of intermittent demand, Croston's method (Croston, 1972) has become the theoretical and practical benchmark for inventory models and forecasts (Teunter, Syntetos, and Babai, 2010). Croston's method differs from previous exponentially smoothed forecasts in that it uses two combined forecasts instead of one forecast based on a single series.

Croston's Method: 50 years of forecasting intermittent demand

The sporadic early research related to estimating demand for slow moving parts is often rooted in predicting usage rates for military spare parts, especially onboard ships (Haber, & Sitgreaves, 1970). Estimating demand for spare parts is a notoriously difficult circumstance due to the erratic nature of the demand of products with little or no usage over time. SES was used for forecasting intermittent demand (Willemain, Smart & Schwarz, 2004) but was seldom accurate.

Other techniques for forecasting intermittent demand have been proposed and most make use of Croston's method as the benchmark to judge performance. The papers by Willemain, Smart, Shockor and DeSautels (1994) and Hua, Zhang, Yang and Tan (2007) are examples. The first proposed bootstrapping for forecasting intermittent demand and the latter utilized autocorrelation for this case. A variation of Croston's method is introduced by Snyder (2002) incorporating bootstrapping into the forecast. Teunter and Duncan (2009) improved Croston's method by considering that an order in a period is triggered by a demand in that period.

Forecasting Intermittent Demand

SES is a commonly used statistical approach to forecast intermittent demand that has been shown to perform adequately on real-world intermittent demand data (Syntetos & Boylan, 2005). Since its publication, Croston's method has been challenged, modified and corrected. Rao (1973) presented a correction to the Croston method. Schultz (1987) considers the effect of selected smoothing constants. Willemain, Smart, Shockor, and DeSautels (1994) challenged the assumptions and tested the method under a variety of assumption violations. Bias related to Croston's method was investigated by Syntetos and Boylan (2001) who then proposed a bias correction factor. Levén and Segerstedt (2004) proposed a procedure based on Croston modeled on the Erlang distribution. However, Boylan and Syntetos (2007) criticized it for using flawed forecast accuracy measures and biased forecasts. Underlying stochastic models related to Croston's method were studied by Shenstone and Hyndman (2005). Boylan and Syntetos (2007) investigated forecast accuracy methods related to Croston's method.

Contextual Text Mining Analysis

Using text mining application, we constructed a word cloud from the abstracts of the papers citing Croston's paper. The associated cloud is helpful in revealing the underlying trends and directions in research investigating Croston's method. Meaningless textual data like "the" and "an" are removed or "stemmed" from the abstracts.

ANALYSIS

Google Scholar identified 934 articles that cite Croston (1972) as a source. When sorted by "relevance" by Google's algorithm, it appears the most influential item related to relevance is the number of citations the article has. For context, the first 50 articles examined had from 120 to 3075 citations. The top 100 articles had at least 50 citations. The top 200 had at least 15 citations. The top 400 had at least 2 citations. The top 540 had at least 1 citation. The remaining articles had no citations. Some were from obscure conference proceedings and some were very new articles. Books were omitted from the analysis since they do not have an abstract. The articles citing Croston (1972) were also analyzed by the year they were published.

Croston's Method: 50 years of forecasting intermittent demand

---

Very little research occurred before 1996. Research interest intensified greatly from 2000 to 2019 peaking at 58 articles in 2019.

Figure 1: Number of Articles Citing Croston (1972) by Year

(omitted for space- available on request)

The following term map resulted from the analysis of weighted terms. At first glance the terms may seem mundane and generic. However, the words are representative of the underlying structure of the research publications using Croston's method. Forecasting demand and intermittent demand are both key terms in this literature, resulting in the emphasis on demand

Figure 2: Word Cloud or Term Map based on Abstracts citing Croston (1972)

which appeared over 1300 times.

A text mining analysis of the abstracts of papers citing Croston yield topics or "clusters" that are derived from groupings of important terms in the abstracts are shown in Table 1.

Table 1: Five Key Word Topic Clusters with 15 Key Terms

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
spare	demand	data	demand	inventory
inventory	forecasting	forecasting	supply	demand
system	method	time	chain	distribution
maintenance	intermittent	series	management	level
management	Intermittent demand	timeseries	supply chain	policy
model	Croston	model	model	control
cost	forecast	forecast	inventory	service
support	smoothing	accuracy	forecasting	stock
systems	performance	sales	neural	method
Inventory_management	spare	forecasts	data	inventory control
planning	results	demand	customer	time
service	time	market	lumpy	parameters
aircraft	Croston method	techniques	approach	process
spare inventory	Demand forecasting	planning	proposed	system
spare demand	exponential	approach	learning	model

Croston's Method: 50 years of forecasting intermittent demand

RESULTS

The first grouping represents articles related to "spare parts inventory models." Papers by Eaves and Kingsman (2004), Porras and Dekker (2008), and Sani and Kingsman (1997) are examples. The study by Eaves and Kingman (2004) models demand using the United Kingdom's Royal Air Force spare parts system to classify parts based on demand patterns and applies the appropriate forecasting technique. The study by Porras and Dekker (2008) examine a variety of demand modeling techniques on real data from an oil refinery to compare reorder point estimation procedures for spare parts. Sani and Kingsman (1997) use data from a typical spare parts depot to evaluate periodic inventory control methods and demand forecasting methods for items with intermittent or low demand.

The second grouping of terms represents the articles that discuss the theoretical basis for Croston's method and is focused on the methods for demand forecasting and the variety of proposed models related to Croston (1972) and measuring the model performance. We are naming this category "Method Focus." Syntetos and Boylan (2001) and Johnston and Boylan (1996) are examples of this category. Syntetos and Boylan (2001) identify a bias in Croston's method and propose an unbiased modification. An extensive simulation is performed to confirm the results. Another example of a "Method Focus" piece is Johnston and Boylan (1996), who identified what level of intermittent demand was required for Croston's method to be an improvement over a method using a single model for an exponentially weighted moving average.

The third grouping represents articles that discuss model applications and forecast accuracy. Many of them use real data to measure model accuracy. The papers by Willemain, Smart and Shockor (1994) and Teunter, Syntetos and Babai (2011) are examples. Willemain, Smart, Shockor, and DeSautels (1994) compare Croston's method to exponential smoothing and test it on real world data from industrial sources and demonstrate its robustness and superiority. Teunter, Syntetos and Babai (2011) explore the behavior of Croston's method when a part with intermittent demand approaches obsolescence. Simulations are used to examine and assess the behavior of these proposed modifications of Croston's method.

The fourth grouping relate to managing demand for spare parts titled "Challenging Intermittent Demand Forecasting." Syntetos and Boylan (2006) examine the stock control performance of different forecasting procedures including simple moving average, single exponential smoothing and Croston's method on 3000 stock keeping units of spare parts. Other papers focus on comparative studies using forecasting methods in which demand for inventory and spare parts are challenging to forecast.

"Inventory Management" is the name for the last grouping and it represents the articles that discuss inventory management. The articles generally look at the accuracy of the methods applied to real data sets. Syntetos, Boylan and Croston (2005) paper is one example of papers in this category. The validity of Croston's method is tested on 3000 real intermittent demand items from the automotive industry.

Croston's Method: 50 years of forecasting intermittent demand

DISCUSSION AND CONCLUSIONS

Research based on Croston's work was minimal for many years. With the exception of Silver (1981) and Williams (1984), very little research was conducted related to Croston (1972) before 1994. Willemain, Smart, Shockor and De Sautels (1994)'s research re-invigorated interest in this area. Shortly after in 2001, Syntetos and Boylan published the first of dozens of papers related to Croston (1972). Syntetos and Boylan and a team of their contemporaries have greatly expanded research on forecasting intermittent demand between 2001 and now. Notably, much of the research in this area has been published in the *Journal of* the *Operational Research Society* and the *International Journal of Production Economics.*

REFERENCES

Boylan, J. E., & Syntetos, A. A. (2007). The accuracy of a modified Croston procedure. *International Journal of Production Economics, 107*(2), 511-517.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Journal of* the *Operational Research Society*, *23*(3), 289-303.

Eaves, A. H., & Kingsman, B. G. (2004). Forecasting for the ordering and stock-holding of spare parts. *Journal of the Operational Research Society, 55*(4), 431-437.

Haber, S. E., & Sitgreaves, R. (1970). A methodology for estimating expected usage of repair parts with application to parts with no usage history. *Naval Research Logistics Quarterly, 17*(4), 535-546.

Hua, Z. S., Zhang, B., Yang, J., & Tan, D. S. (2007). A new approach of forecasting intermittent demand for spare parts inventories in the process industries. Journal of the *Operational Research Society, 58*(1), 52-61.

Johnston, F. R., & Boylan, J. E. (1996). Forecasting for items with intermittent demand. *Journal of the Operational Research Society, 47*(1), 113-121.

Levén, E., & Segerstedt, A. (2004). Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics, 90*(3), 361-367.

Porras, E., & Dekker, R. (2008). An inventory control system for spare parts at a refinery: An empirical comparison of different re-order point methods. *European Journal of Operational Research, 184*(1), 101-132.

Rao, A. V. (1973). A comment on: forecasting and stock control for intermittent demands. Journal of the *Operational Research Society, 24*(4), 639-640.

Sani, B., & Kingsman, B. G. (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society 48*(7), 700-713.

Schultz, C. R. (1987). Forecasting and inventory control for sporadic demand under periodic review. Journal of the *Operational Research Society, 38*(5), 453-458.

Shenstone, L., & Hyndman, R. J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting, 24*(6), 389-402.

Silver, E. A. (1981). Operations research in inventory management: A review and critique. *Operations Research, 29*(4), 628-645.

Snyder, R. (2002). Forecasting sales of slow- and fast-moving inventories. *European Journal of Operational Research, 140*(3), 684-699.

Syntetos, A. A., & Boylan, J. E. (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics, 71*(1-3), 457-466.

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of forecasting, 21*(2), 303-314.

Syntetos, A. A., & Boylan, J. E. (2006). On the stock control performance of intermittent demand estimators. *International Journal of Production Economics, 103*(1), 36-47.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society, 60, 56*(5), 495-503.

Teunter, R. H., & Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society, 60*(3), 321-329.

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2010). Determining order-up-to levels under periodic review for compound binomial (intermittent) demand. *European Journal of Operational Research, 203*(3), 619-624.

Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research, 214*(3), 606-615.

Willemain, T. R., Smart, C. N., Shockor, J. H., & DeSautels, P. A. (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International journal of forecasting, 10*(4), 529-538.

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of forecasting, 20*(3), 375-387.

Williams, T. M. (1984). Special products and uncertainty in production/inventory systems. *European Journal of Operational Research, 15*(1), 46-54.

DECISION SCIENCES INSTITUTE

A hybrid approach for scheduling jobs on identical parallel machines

Pravin P. Tambe
Indian Institute of Management Tiruchirappalli
Email: pravin@iimtrichy.ac.in

ABSTRACT

Parallel machine manufacturing is a common setup that widely exists in industries. An identical parallel machine scheduling problem for minimizing total tardiness is a very important production scheduling problem, but there have been many difficulties in solving large-scale problems. Metaheuristic approach like Simulated Annealing has shown efficient results in solving the combinatorial optimization problem. In this paper, a hybrid approach of a simulated annealing algorithm combined with a backward-forward heuristic has been proposed for solving identical machine scheduling problem for minimizing the total tardiness. A numerical example of scheduling jobs on identical high-pressure die casting machines is presented.

KEYWORDS: Scheduling, Parallel Machine Problem, Simulated Annealing, Hybrid Approach

INTRODUCTION

Scheduling is one of the most critical functions in any manufacturing organization due to limited resources, increased customer expectations, and fierce competition both domestically and internationally. Meanwhile, cost reduction and profit maximization continue to be strong motivations for all manufacturing companies in the present global environment. As tardiness relates to operational costs, manufacturing schedules can affect the performance of a company and hence chance of its survivability. The parallel machine scheduling problem is one of the most popular scheduling problems due to the interest of both the academic and industrial community.
Scheduling jobs on parallel machines or processing units for meeting due dates is a common situation found in many applications. For example, in case of a bakery manufacturing variety of loaves of bread and bakery products, which needs to be baked in ovens having the same capacity, in the printing industry, where jobs to be printed are scheduled on parallel printing presses and other finishing tasks. In this case, meeting due dates for orders placed by the different book publishers is extremely important. In the machine shop industry, different jobs are to be assigned to different machines when certain types of jobs require more processing time and the batch quantity of the particular jobs is high, then it is required to provide more than one machine of a similar type to complete the jobs in time. Similarly, the pharmaceutical industry usually runs parallel machines to produce drugs, rolling mills for steel products, etc.
This paper addresses the application of a hybrid approach to parallel machine earliness-tardiness non-common due date scheduling problem for independent jobs with varying processing times. For the earliness penalty, the inventory holding cost for the batches is considered. The numerical example of scheduling batches on three identical parallel machines is presented. The computational experiment for different problem sizes and schedule tightness factors is also presented.

LITERATURE REVIEW

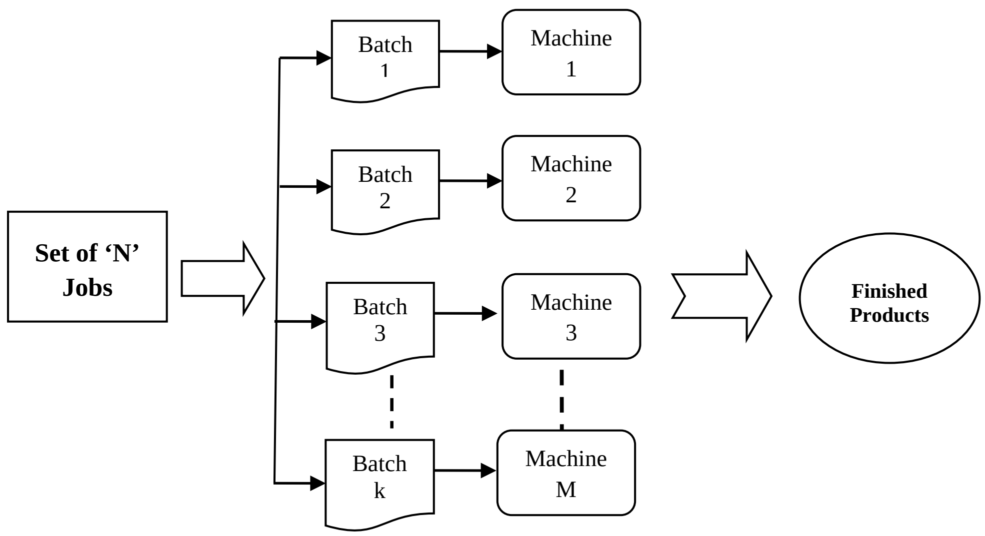
The parallel machine environment has been studied for several years because of its importance to academia and industry. In literature, several optimizing criteria and objectives are defined to determine the most efficient and effective parallel machine schedule. Minimizing total tardiness is the most interesting criterion for production systems. Koulamas (1997) presented a polynomial decomposition of the identical parallel machine environment with a total tardiness objective function in which meta-heuristic simulated annealing was employed. Min and Cheng (1999) study on the application of GA in solving identical parallel machine scheduling problem for minimizing the makespan and developed machine-code based genetic algorithm method which performs better for large scale scheduling problems. Radhakrishnan and Ventura (2000) studied the problem of parallel machine earliness- tardiness, non-common due date sequence-dependent set-up time scheduling problem (PETNDDSP) for jobs with varying processing times. The objective is to minimize the sum of the absolute deviations of job completion times from their corresponding due dates. The computational study shows that using the SA methodology, significant improvements to the local search heuristic solutions can be achieved for such problems. Yalaoui and Chu (2002) present a branch and bound algorithm to obtain optimal solutions in identical parallel-machine scheduling problems to minimize total tardiness. Liaw et al. (2003) addressed the problem of minimizing the total weighted tardiness of unrelated parallel machines, showed the properties of an optimal schedule, and also proposed a branch and bound algorithm. Cao et al. (2005) addressed the problem of simultaneously selecting and scheduling parallel machines to minimize machine costs and tardiness cost using a Tabu metaheuristic approach. Bilge et al. (2007) applied a Tabu search to a single machine, total weighted tardiness problem with due dates. Lee et al. (2006) presented a makespan minimization scheduling problem on identical parallel machines using a simulated annealing method. Computational results demonstrated that the proposed heuristic is very accurate and that it outperforms the existing methods. Biskup et al. (2008) addressed minimization of total tardiness for an identical parallel machine problem and discussed some heuristics such as the traffic priority index (TPI) heuristic, the modified due date (MDD) heuristic, and KPM. New solution approaches were developed based on these heuristics. Lin et al. (2011) proposed a GA for unrelated parallel machine problem to minimize the total weighted tardiness. Lin et al. (2013) proposed ant colony optimization incorporating some new ideas such as heuristic initial solution, machine reselection step, and local search procedure to solve the problem of scheduling unrelated parallel machines to minimize total weighted tardiness. Kerkhove and Vanhoucke (2014) presented a case study at a Belgian textile manufacturer whose production lines are situated in dispersed locations, each of which containing multiple machines operating in parallel. Hessler and Deghdak (2017) proposed the discrete parallel machine makespan (DPMM) Scheduling- Location(ScheLoc) problem that differs from the previous versions by considering ‘p’ parallel machines whose locations should be selected from a discrete set of candidate locations. Majumder et al. (2018) dealt with the problem of scheduling identical parallel batch processing machines where each machine processes a set of jobs in a batch simultaneously and each job in the batch is characterized by its processing time, ready time, and job size. They proposed a hybrid discrete algorithm based on a modified variable neighborhood search and cuckoo search algorithm to minimize makespan for the scheduling problem. Báez et al. (2019) proposed a hybrid algorithm that combines GRASP and Variable Neighborhood Search for solving a parallel machine scheduling problem with dependent setup times for minimizing the total completion time. The objective is to assign jobs to machines and determine the order in which jobs have to be processed on the machines, in such a way that the sum of the completion times of all jobs is minimized. Croce et al. (2021) considered the parallel machine scheduling problems involving the minimization of the weighted or unweighted number of tardy jobs. They proposed exponential-time approximation algorithms and fixed-parameter tractable exact algorithms to solve them. Mönch and Shen (2021) addressed a parallel machine scheduling problem in a distributed manufacturing setting. The objective is to minimize the total weighted delivery time (TWD) after including delivery durations for the jobs. They proposed a greedy randomized adaptive search (GRASP) framework. Kim and Lee (2021) considered a uniform parallel machine scheduling problem with machine eligibility, job splitting, sequence-dependent setup times, and limited setup servers, which has been defined from a real application of piston manufacturing in Korea. The objective is to minimize the makespan. They developed an efficient heuristic algorithm that assigns jobs to machines by considering processing times, setup times, and machine eligibility constraints, and then balances machine workloads by splitting jobs.

PROBLEM DESCRIPTION

Most of the time, real-life scheduling problems involve multiple machines. Multiple machines may occur in series, parallel, or both. In parallel machine scheduling, each job can be processed on any of the identical machines and processing times are independent of the machine. The advantage with this type of manufacturing system is the considerable reduction in makespan. The parallel machine manufacturing system is shown in Figure1. The scheduling decisions involved in parallel machine problem are:

i) Which machine processes each job?
ii) In what sequence the jobs to be processed on the machine.

Figure 1: Parallel machine scheduling problem



The production unit under study consists of M identical machines producing different product types. A batch scheduling problem where a set of identical parallel machines is available for processing the jobs in batches in the continuous mode is addressed. Formally, there is a set of jobs available from time zero onwards for processing on M identical parallel machines in batches. Each batch ‘k’ has a processing time and a due date; before which the batch is expected to

complete. The batches are to be grouped for processing with a sequence-independent setup time; The jobs in the same batch are continuously processed after the setup operation and the processing length of a batch is the sum of the setup time and the processing times.
Following assumptions are made for the batch scheduling problem.

1. Each machine has the ability for processing each job.
2. The batch is available at time zero.
3. The machine can process no more than one batch at a time.
4. Each batch is independent of the other.
5. A batch cannot be pre-empted by another batch.
6. The setup time is included in the processing time for the job.

Let P_k denote the processing time of the batch 'k', let DD_k denote the due date; before which the batch is expected to complete, and W_k be the penalty cost for the batch k, if it is late. Also, let CT_k denote the completion time of the k^{th} batch, where k=1,2,……….n, and 'n' is the number of jobs to be processed on the j^{th} machine; j=1,2 ………M.
The lateness of the job can be defined as,

$$\mathrm{LT_k} = CT_k - DD_k \tag{1}$$

and the tardiness penalty of the k^{th} job is,

$$T_k = W_k \times \{max[0, LT_k]\} \tag{2}$$

Let $'ST'_k$ denote the start time of the k^{th} batch in the production schedule. Then, the inventory holding cost is calculated as follows:

i) If $\mathrm{DD_k} \geq \mathrm{CT_k}$, then

$$\mathrm{IHC_k} = \left\{ \int_0^{(\mathrm{CT-ST})} (\mathrm{CT - ST - t}) \times \mathrm{PR} \times \mathrm{Ch} \times \mathrm{dt} + (\mathrm{DD - CT}) \times \mathrm{PR} \times \mathrm{Ch} \times (\mathrm{CT - ST}) \right\}_\mathrm{k} \tag{3}$$

ii) If $DD_k < ST_k < CT_k$

$$\mathrm{IHC_k} = \left\{ \int_0^{(CT-ST)} (CT - ST - t) \times PR \times Ch \times dt \right\}_k \tag{4}$$

iii) If $ST_k < DD_k < CT_k$

$$\mathrm{IHC_k} = \left\{ \int_0^{(DD-ST)} (DD - ST - t) \times PR \times Ch \times dt + \int_0^{(CT-DD)} (CT - DD - t) \times PR \times Ch \times dt \right\}_k \tag{5}$$

The objective of production scheduling is to find an optimal sequence of batch such that the total cost of batch delay and inventory holding cost for all jobs is minimized. The total schedule penalty cost is given by,

$$\text{Schedule Penalty Cost} = \sum_{j=1}^{M} \sum_{\mathrm{k=1}}^{\mathrm{n}} \mathrm{W_k} \times (\max[0, \mathrm{LT_k}]) + \mathrm{IHC_k} \tag{6}$$

where 'n' is the number of batches on the allocated machine. The objective function is to minimize the total penalty cost.
The next section presents the simulated annealing algorithm.

SIMULATED ANNEALING

Simulated Annealing (SA) is an iterative search method proposed by Kirkpatrick et al. (1983). It is a generalized probabilistic approach for approximately solving large combinatorial optimization problems. The simulated annealing method is based on the simulation of thermal annealing of critically heated solids. When a solid (metal) is brought into a molten state by heating it to a high temperature, the atoms in the molten metal move freely with respect to each other. However, the movements of atoms get restricted as the temperature is reduced. As the temperature reduces, the atoms tend to get ordered and finally form crystals having the minimum possible internal energy. The process of formation of crystals essentially depends on the cooling rate.

This physical behaviour of the annealing process is simulated in the SA technique to find the optimal or near-optimal solutions for complex combinatorial optimization problems. When applied to optimization problems, the SA algorithm procedure begins with an initial feasible solution that is randomly generated. During the search process, the algorithm generates a new solution based on some perturbation mechanism, in the neighborhood of the current solution. If the newly generated solution is better than the current solution, then the newly generated solution replaces the current solution. If the new solution is inferior to the current solution, the algorithm will accept the inferior solution using Metropolis's criterion, which is based on Boltzmann's probability. According to Metropolis's criterion, if the difference between the cost function values of the current and the newly generated solutions (ΔG) is equal to or larger than zero, a random number R_n in [0,1] is generated from a uniform distribution. If $R_n < \exp(-\Delta G/T_p)$, the newly generated solution is accepted as the current solution.

The acceptance of inferior solutions with a certain probability helps the algorithm to avoid being trapped at a local optimum. Finally, the algorithm terminates using a stopping criterion. In literature, several criteria have been used by the researchers for stopping the iterative procedure like; the maximum number of iterations, minimum temperature value, minimum value of the objective function, or minimum value of acceptance rate.

The SA procedure consists of four basic components (Ghodratnama et al., 2010),

i) Configuration defines all of the possible solutions for the combinatorial problem.

ii) The move set defines the set of allowable transitions. These transitions must be capable of reaching all of the configurations.

iii) The cost function defines a measure of how good any given configuration.

iv) The cooling schedule defines the annealing of the problem from a random to a good, frozen solution. It is worth noting that the cooling schedule determines the initial temperature, the rule for decreasing the value of temperature, the number of iterations for searching for better configurations at each temperature, and the time at which annealing should be stopped.

In general, the algorithm may be summarized into the following steps:

Step 1: Choose an initial state randomly and assign an initial temperature.

Step 2: Generate a new state over sample size at temperature T.

Step 3: Calculate the new state energy.

Step 4: Compare the difference between the energy of the new state and the old state. If the new state energy is less than, the previous one, then accept the new state, else accept the new state only if it satisfies a certain probability.

Step 5: Decrease the temperature.

Step 6: If the energy is less than a certain value then the initial state of the system may be generated randomly or selected using any heuristic method. A proper choice of initial temperature is a sufficiently high value to accept the first few selected states. There is a lot of flexibility in choosing a generation scheme. The acceptance probability decreases as temperature decreases. The general approach of the algorithm which is used for optimization is shown in Figure 2.

Figure 2: Simulated annealing algorithm

Select
An initial temperature, T_p (a large number)
An initial solution, S_o
A cost function, G
A temperature reduction factor, θ_S
A neighborhood structure for the solution space
Repeat
Repeat
new solution, S_{New} =perturb(S_o)
$\partial G = G(S_{New}) - G(S_o)$
If $\partial G \leq 0$ **OR** random number $R_n[0, 1] < \exp(-\partial G/T_p)$
then, $S_o = S_{New}$
Until iteration count = Max_number_iteration
$T_p = T_p \times \theta_S$
Until stopping condition

BACKWARD-FORWARD HEURISTIC

The backward-forward heuristic (Sule, 2007) works in two phases. The first phase, i.e. backward phase is similar to obtaining the first feasible solution and the second phase, i.e. forward phase involves swapping the job positions in a systematic manner and a new solution is accepted only if it leads to a lower penalty. The steps of the heuristic are as follows:
A) Backward phase
In the backward phase, an initial batch sequence is developed. The sequential batch assignments start from the last position and proceed backward towards the first position. The assignments are complete when the first position is assigned a batch. The steps of the backward phase are as follows:
i) Note the position in the sequence (the value of the position counter) in which the next batch is to be assigned. The sequence is developed starting from position ‘m’ and continuing backward to position 1. The initial value of the position counter is ‘m’.
ii) Calculate the sum of processing times for all unscheduled batches.
iii) Calculate the penalty for each unscheduled batch.
iv) The next batch to be scheduled in the designated position (the value of the position counter) is the one having a minimum penalty from step iii. In case of a tie, choose the batch with the largest processing time.
v) Reduce the position counter by 1.
Repeat steps (i) through (v) until all the batches are scheduled.
B) Forward phase
The batch sequence obtained in the backward phase is the “best” sequence at this stage. The forward phase is to be performed on the sequence obtained from the backward phase for further improvement. The forward pass progresses from the batch position ‘1’ toward the batch in position ‘m’. Let K_{lag} define the lag between two batches in the sequence that are to be exchanged (for example, the batches in positions 1 and 3 have lag, K_{lag}=2). The steps of the forward phase are as follows:
i) Set K_{lag}= m-1.
ii) Set J= K_{lag} +1.

iii) Determine the saving (or cost) by exchanging two batches in the “best” sequence with a lag of $K_{lag.}$ The batch scheduled in position J is exchanged with the batch scheduled in position J-K_{lag} (if J-K_{lag} is zero or negative, go to step vi). Calculate the penalty after the exchange, and compare it to the “best” sequence penalty.
iv) If there is either positive or zero savings in step iii, go to step v; otherwise, there is cost associated with this exchange and the exchange is rejected. Increase the value of J by 1. If J is equal to or less than m, go to step iii. If J is greater than m go to step vi.
v) If the total penalty has decreased, the exchange is acceptable. Perform this exchange. The new sequence is now the best sequence; go to step i. even if the saving is zero, make the exchange and go to step i, unless the set of batches associated in this exchange has been checked and exchanged in an earlier application of the forward phase. In that case, no exchange is made at this time. Increase the value of J by 1. If J is less than m, go to step iii. If J=m, go to step vi.
vi) Decrease the value of K_{lag} by 1. If K_{lag}>0, go to step ii. If K_{lag} =0, go to step vii.
vii) The resulting sequence is the best sequence generated by this procedure.

NUMERICAL EXAMPLE

In this section, the numerical example of scheduling jobs in the aluminum die casting industry is presented. The case of a particular order of 30 batches to be processed on three identical, high-pressure die casting machines is considered. The machines under consideration operate for three shifts a day for eight hours to carry out the production. The details of the production parameters such as average batch size, processing times, due date, etc. for the production order are given in Table 1.

Table 1: Production parameters for the three-machine problem.

Batch	Penalty Cost (Rs. per hour)	Profit Per Unit (Rs./uni)	Cost per Item (Rs./unit)	Batch Size	Processing Time (hr.)	Inventory Holding Cost (Rs./unit / hr.)	Due Date (hr.)
1	660	22	147	7050	47	0.0034	424
2	1200	20	132	8400	56	0.0030	505
3	383	17	113	14100	94	0.0026	847
4	360	12	77	13650	91	0.0018	820
5	360	12	83	5700	38	0.0019	343
6	420	14	96	6150	41	0.0022	370
7	405	18	121	4650	31	0.0028	280
8	240	8	53	6900	46	0.0012	415
9	330	11	70	3450	23	0.0016	208
10	660	22	149	10500	70	0.0034	631
11	1140	19	127	11700	78	0.0029	703
12	570	19	129	13200	88	0.0029	793
13	480	16	105	9000	60	0.0024	541
14	1320	22	148	9750	65	0.0034	586
15	840	14	91	6450	43	0.0021	388

16	203	9	61	11100	74	0.0014	667
17	360	16	107	8850	59	0.0024	532
18	960	16	109	9450	63	0.0025	568
19	270	12	82	9900	66	0.0019	595
20	240	8	50	12600	84	0.0011	757
21	600	10	64	11550	77	0.0015	694
22	840	14	95	8700	58	0.0022	523
23	203	9	61	5700	38	0.0014	343
24	338	15	97	9900	66	0.0022	595
25	270	9	60	11250	75	0.0014	676
26	540	9	62	7050	47	0.0014	424
27	248	11	72	5700	38	0.0016	343
28	248	11	71	10800	72	0.0016	649
29	420	14	94	9750	65	0.0021	586
30	540	9	58	12000	80	0.0013	721

The next sub-section presents the application of the proposed approach to obtain the near-optimal schedule for the batches.

Obtaining the optimal production schedule

The problem in the present case is to schedule 30 batches on 3 identical parallel machines to minimize the total tardiness penalty cost. Each job's processing rate, due date, and penalty cost are given in Table 1. As already discussed, a simulated annealing combined with a backward-forward heuristic is used for scheduling. Both earliness and tardiness penalties are considered in the problem. The sum of earliness and tardiness penalty cost of the schedules on each machine is calculated and the total penalty cost for all the machine is used as the acceptance criteria during the iterations of the SA algorithm. The solution to the problem is to allocate the batches on each machine and sequence the batches such that the penalty cost for the assigned batch on the machine is minimum. For implementation, the important steps of the SA algorithm mentioned in Figure 2 are explained below.

<u>Initial solution</u>

The first step in the implementation of the SA is to generate the initial solution. Generally, the initial solution is randomly generated and jobs are randomly assigned to machines or using some heuristic rules like shortest processing time (SPT) or the earliest due date (EDD). In the present case, for an initial solution, we consider a new approach to assign the batches to each machine. The sum of processing times for all the batches is calculated and divided by the number of machines to get the processing time per machine which is denoted as PTM. A sequence for all the jobs is generated randomly and then starting from the first batch, the batches are assigned one by one to machines such that the total time for the assigned batches on each machine is nearly equal to the PTM. The batches assigned on each machine are scheduled using the backward forward heuristic (Sule, 2007) to get the minimum tardiness penalty cost sequence on each machine. This will result in a better feasible solution as compared to the random allocation. If the jobs are randomly assigned to the machines, then no attention will be paid to the total time and the tardiness penalties. Hence, the initial feasible solution generated may be far away from

optimality. In the present example, there are 30 jobs to be scheduled and the PTM is equal to 611. As an initial solution, the batches allocation to each machine having a total processing time nearly equal to PTM is given below:
Machine 1: 28 6 9 30 7 3 22 16 17 11
Machine 2: 27 21 15 5 1 19 8 26 14 29 25
Machine 3: 12 23 20 2 24 13 4 18 10

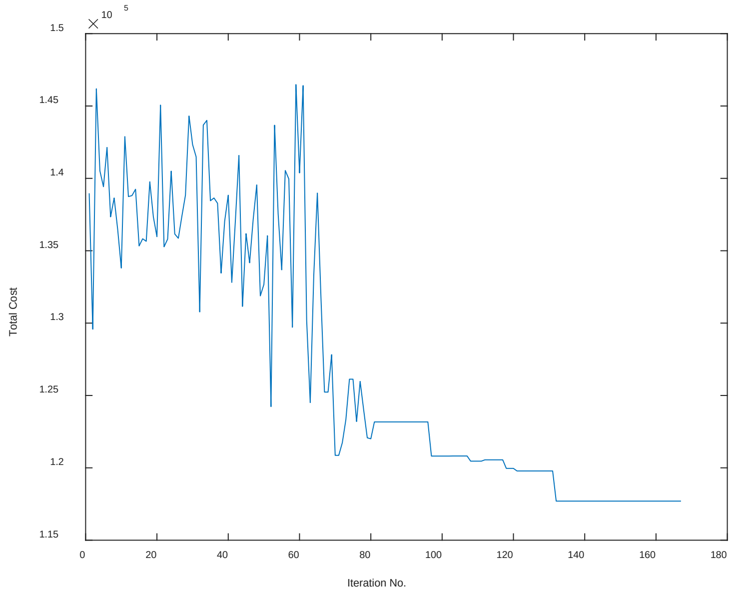
Neighborhood generation

The neighborhood structure plays an important role in the efficiency of the SA method. A new neighboring solution is generated by making a small change in the current assignment of jobs and then evaluating this new solution. In order to generate a neighboring solution, a swap rule is implemented. The machines are selected randomly and then the batches are swapped on the selected machines to generate a new solution. The batches are then sequenced using the backward-forward heuristic and the schedule penalty cost is calculated for the batch sequence on each machine.
The parameters selected for the implementation of the SA algorithm to production scheduling problem are an initial temperature (T_p) of 1000000, a temperature reduction factor of 0.92, and termination criteria of T_p<1. The algorithm is implemented using a program developed in Matlab R2017a and implemented on a computer with a 2.40GHz processor and 8GB RAM. The near-optimal schedule obtained on the machine 1, 2 and 3 are given in Table 2 and the total cost progress during the iterations of SA algorithm is given in Figure 3.

Table 2: Near-Optimal batch sequence on machines

Machine No.	Sequence of batches	Penalty Cost (Rs.)
Machine 1	27- 25- 14 - 23 - 7 - 22 - 24 - 2 - 29	37040
Machine 2	10- 4 - 1 - 8 - 6 - 11 - 30 - 13 - 19 - 28	33968
Machine 3	18 - 21 - 9 - 5 - 12 - 5 - 26 - 3 - 17 - 16 - 20	46692

Figure 3: Iterations of SA algorithm



COMPUTATIONAL EXPERIMENT

A computational experiment is performed to get the results for scheduling batches on 3 machines with different problem sizes and different levels of the schedule tightness factor (Sule, 2007) of loose (0.4), moderate (0.6), and tight (0.7). The production-related parameters are generated randomly similar to the above problem and assuming a uniform distribution for the data values. For example, the processing times are generated randomly using U(20,80). Similarly, the other parameters are generated randomly for the problems. The results of the experiment are given in Table 3. The results indicate that the computation time is increasing with the increase in problem size. Also, for a large size problem, the tight schedule requires more time as compared to loose and moderate tightness factor.

Table 3: Results for a 3-machine scheduling problem

Problem Batch Size	Tightness factor	Total Cost	Computation Time (Sec.)	Allocation of batches to machines
	Loose	8836	37.14	[4 2 3]
9	Moderate	12754	24.24	[3 3 3]
	Tight	4992	34.94	[4 2 3]
	Loose	29253	201.13	[4 6 5]
15	Moderate	33900	194.49	[6 5 4]
	Tight	10576	201.57	[6 5 4]

20	Loose	62729	668.42	[6 6 8]
	Moderate	42455	1011.61	[9 5 6]
	Tight	29622	784.72	[7 5 8]
25	Loose	79813	2297.13	[6 11 8]
	Moderate	69605	1744.11	[7 9 9]
	Tight	46205	1740.77	[8 8 9]
30	Loose	122230	10533.24	[8 11 11]
	Moderate	100694	10534.20	[11 10 9]
	Tight	90745	10935.58	[9 9 12]
40	Loose	244552	12310.29	[14 14 12]
	Moderate	195922	11778.04	[13 13 14]
	Tight	182315	11972.56	[13 13 14]
50	Loose	403367	33256.77	[18 16 16]
	Moderate	251299	33443.67	[17 18 15]
	Tight	281521	36326.58	[16 16 18]

CONCLUSION

In this paper, a hybrid approach of simulated annealing combined with a backward-forward heuristic is presented to solve the identical parallel machine problem. The objective is to find a schedule that minimizes the total penalty cost while simultaneously determining the assignment of jobs to available machines and sequencing the assigned jobs on each machine. The schedule penalty cost includes earliness as well as tardiness penalty cost. In future, the computational performance of the proposed hybrid approach can be compared with other approaches like a branch and bound, and heuristic/meta-heuristic approaches. Also, the problem complexity like job splitting, sequence-dependent setup times, etc. needs to be explored. The scheduling of jobs in a parallel machine environment without considering maintenance increases the probability of machine breakdowns, and inversely, considering this action in production can prolong the completion time of jobs. Hence, future work can be done to combine the scheduling of parallel machines with maintenance for simultaneously optimizing the scheduling and maintenance decision on the machines.

REFERENCES

Báez, S.,Angel-Bello,F.,bAlvarez,A. and Melián-Batista, B. (2019). A hybrid metaheuristic algorithm for a parallel machine scheduling problem with dependent setup times, *Computers & Industrial Engineering*,131: 295-305

Bilge Ü, Kurtulan M and Klraç F. (2007). A tabu search algorithm for the single machine total weighted tardiness problem, *European Journal of Operational Research*, 176:1423-1435.

Biskup D., Herrmann J. and Gupta J.N.D. (2008). Scheduling identical parallel machines to minimize total tardiness, *International Journal of Production Economics*, 115(1):134-142.

Cao D., Chen M., Wan G. (2005). Parallel machine selection and job scheduling to minimize machine cost and job tardiness, *Computers and Operation Research*, 32:1995-2012.

Croce,F., T'kindt,V. and Ploton, O. (2021). Parallel machine scheduling with minimum number of tardy jobs: Approximation and exponential algorithms, *Applied Mathematics and Computation*, 397:125888

Ghodratnama A., Rabbani M., Tavakkoli-Moghaddam R, Baboli A. (2010). Solving a single-machine scheduling problem with maintenance, job deterioration and learning effect by simulated annealing, *Journal of Manufacturing Systems*, 29:1–9

Hessler, C., Deghdak, K. (2017). Discrete parallel machine makespan ScheLoc problem, *Journal of Combinatorial Optimization*, 34 (4):1159–1186.

Kerkhove, L.P. and Vanhoucke, M., (2014). Scheduling of unrelated parallel machines with limited server availability on multiple production locations: a case study in knitted fabrics. *International Journal of Production Research*, 52 (9):2630–2653.

Kim, H. and Lee, J. (2021). Scheduling uniform parallel dedicated machines with job splitting, sequence-dependent setup times, and multiple servers, *Computers & Operations Research*, Volume 126: 105115

Kirkpatrick S., Gelatt C.D. and Vecchi M.P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598): 671-680.

Koulamas, C. (1997), Decomposition and hybrid simulated annealing heuristics for the parallel-machine total tardiness problem, *Naval Research Logistics*, 44:109-125.

Lee, W., Wu, C. and Chen P. (2006). A simulated annealing approach to makespan minimization on identical parallel machines, *International Journal of Advanced Manufacturing Technology*, 31:328–334

Liaw C-F, Lin Y-K, Cheng C-Y, Chen M (2003). Scheduling unrelated parallel machines to minimize total weighted tardiness. *Computers and Operation Research*, 30:1777-1789.

Lin, C. W., Lin, Y. K. and Hsieh, H. T. (2013). Ant colony optimization for unrelated parallel machine scheduling. *The International Journal of Advanced Manufacturing Technology*, 67: 35-45.

Lin, Y. K., Pfund, M. E.,and Fowler, J. W. (2011). Heuristics for minimizing regular performance measures in unrelated parallel machine scheduling problem, *Computers & Operations Research*, 38(6):901–916

Majumder, A., Laha, D. and Suganthan, P.N. (2018). A hybrid cuckoo search algorithm in parallel batch processing machines with unequal job ready times, *Computers & Industrial Engineering*,124:65-76

Min L. and Cheng W. (1999). A genetic algorithm for minimizing the makespan in the case of scheduling identical parallel machines, *Artificial Intelligence in Engineering*, 13:399–403.

Mönch, L., Shen, L. (2021). Parallel machine scheduling with the total weighted delivery time performance measure in distributed manufacturing, *Computers & Operations Research*, me 127:105126

Radhakrishnan S. and Ventura J. (2000), Simulated annealing for parallel machine scheduling with earliness-tardiness penalties and sequence dependent set-up times, *International Journal of Production Research*, 38(10):2233-2252

Sule, D.R. (2007). Production Planning and Industrial Scheduling, CRC Press, Boca Raton.

Yalaoui, F.and Chu, C. (2002). Parallel machine scheduling to minimize total tardiness. *International Journal of Production Economics,* 76:265–279.

DECISION SCIENCES INSTITUTE

A branch-and-bound algorithm for minimizing weighted squared tardiness on a single machine with release dates

Jeffrey Schaller
Department of Business Administration
Eastern Connecticut State University
83 Windham St.
Willimantic, CT 06226 -2295
(860) 465 - 5226
FAX: (860) 465 - 54469
schallerj@easternct.edu

Jorge M. S. Valente
Faculdade de Economia,
Universidade do Porto
Rua Dr. Roberto Frias, 4200-464 Porto
Portugal
jvalente@fep.up.pt

James T. L. Sambuchetti
Faculdade de Economia
Universidade do Porto
Rua Dr. Roberto Frias, s/n, 4200-464 Porto
Portugal

ABSTRACT

This paper considers a problem in which there is a set of jobs to be sequenced on a single machine. Each job has a weight, a due date, a processing time and a release date. The objective is to sequence the jobs to minimize total weighted squared tardiness. A branch-and-bound algorithm is developed for optimally solving the problem. The algorithm includes several conditions for curtailing the algorithm's tree and a lower bound for each node. The algorithm is shown to be able to solve instances 10 and 15 jobs in a reasonable amount of time.

KEYWORDS: Scheduling, Branch and Bound, Single Machine, Weighted Squared Tardiness

INTRODUCTION

Meeting customer due dates is a very important element in providing excellent customer service and retaining customers. Many customers are trying eliminate as much waste as possible from their supply chains and have come to view tardy delivery of products as undesirable. In some cases, tardy deliveries can result in lost sales and in other cases may cause loss of customer goodwill and possibly result in losing a customer. Modelling the cost of tardiness can be difficult but one approach is to use a quadratic penalty to represent a customer's dissatisfaction with the tardiness. This approach is derived from the loss function proposed by (Taguchi, 1986) for measuring quality. This paper considers the objective of minimizing the sum of weighted squared tardiness values for all jobs to be processed on a single machine when jobs have distinct due dates and release times.

We describe the problem formally as follows. A set $N = \{1,2, \ldots, n\}$ of n independent jobs is to be processed on a single machine. This machine is assumed to be always available, and it can only work on one job at a time. Also, pre-emption is not allowed, so once the machine begins working on a job, that job must be processed without interruption.

Each job $j, j \in N$ cannot start processing before it is available which is denoted by its release date r_j. Also, each job $j, j \in N$ requires a processing time p_j, and has a weight w_j and a due date d_j. The completion time of job $j, j \in N$ will be denoted by C_j. For any given sequence of jobs, the tardiness of job $j, j \in N$, is defined as $T_j = max\{C_j - d_j; 0\}$. The objective is to determine a sequence of the jobs that minimizes the total weighted quadratic tardiness $\sum_{j=1}^{n} w_j T_j^2$.

Finding the permutation sequence of jobs that minimizes the objective is challenging. One approach is to exhaustively enumerate all the combinations of sequences and pick the one with the lowest objective. A second approach is to use implicit enumeration approach, such as branch-and-bound to find an optimal sequence. A third approach is to use heuristic methods that find reasonably good solutions quickly. The first two approaches result in an optimal sequence being found but as the size, in terms of numbers of jobs to be sequenced, increases the time needed can become quickly excessive. The third approach is likely to be able to generate solutions for large scale instances but the quality of the solutions may suffer.

Our focus in this research is on the second approach, the development of branch-and-bound algorithms that generate an optimal solution. The use of branch-and-bound should be more efficient than exhaustive enumeration and allow for the finding of optimal solutions for larger instances than the former. We plan to develop heuristic algorithms that can generate solutions for larger sized instances but want to have an algorithm that can generate optimal solution for smaller size instances that can then be used as a benchmark to compare the accuracy of solutions generated by heuristic algorithms.

LITERATURE REVIEW

In this paper the single machine environment is considered. We chose this environment for several reasons. First, scheduling in a single machine environment occurs in several practical operations; for a specific example in the chemical industry, see (Wagner, Davis and Kher, 2002). Second, in many production systems there is a single bottleneck machine and scheduling this single machine determines the performance of the system. Third, by studying single machine problems, results and insights can be obtained that can often be applied to more complex scheduling environments, such as flow shops or job shops. The objective considered in this research is the sum of weighted squared tardiness so we focus the review of the literature on problems with this objective, particularly for a single machine.

There been relatively few papers that have considered the objective of minimizing the sum of weighted squared tardiness values. Several of the approaches for the weighted squared tardiness objective use a Lagrangian relaxation in which the machine capacity constraints are relaxed to obtain a lower bound and then a heuristic is used to create a feasible solution and obtain an upper bound. The Lagrangian relaxation and its solution procedure is based on the one used by (Fisher, 1973a and 1973b) for other objectives. (Sun et al., 1999) considered the single machine problem with release dates and sequence dependent setup times. They compared their Lagrangian relaxation-based heuristic against some simple dispatching rules, a tabu search and simulated annealing algorithms. These heuristics were tested using a variety of data sets most of which consisted of 40 jobs and ranged between 10 and 80 jobs. (Hoitomt et al., 1990) and (Luh and Hoitomt, 1993) developed procedures for scheduling jobs on parallel machines. (Hoitomt et al., 1990)'s procedure was for parallel machines in which jobs have multiple operations with precedence constraints. The procedure is demonstrated on three examples from a Pratt and Whitney plant. (Luh and Hoitomt, 1993)'s procedure was for identical parallel machines and was also demonstrated using data from a Pratt and Whiney plant including an example with 112 jobs and 44 machines. This environment was also considered by (Schaller and Valente, 2018). They analysed several efficient dispatching rules, proposed an improvement procedure and presented Lagrangian relaxation methods and an ILS metaheuristic. (Luh and Hoitomt, 1993), (Sun and Noble, 1999) and (Thomalla, 2001) considered the job shop scheduling problem. (Sun and Noble, 1999) considered the job shop scheduling problem with sequence dependent setups and (Thomalla, 2001) considers the problem with alternative processing plans.

(Schaller and Valente, 2012) developed a branch-and-bound algorithm that included several dominance rules for minimizing total weighted squared tardiness in a single machine without release dates. Multiple efficient dispatching procedures were analysed in (Valente and Schaller, 2012). (Gonçalves, Valente et al., 2016) proposed a local search procedure, and incorporated it in three metaheuristics, namely iterated local search (ILS), variable greedy (VG) and steady-state genetic algorithm (SSGA).

(Costa, Valente et al., 2020) addressed the problem of minimizing total weighted squared tardiness in flow shops. They analysed several efficient dispatching rules, which can quickly solve even quite large instances. Various improvement procedures were also considered, and these were applied to the best dispatching rule. The non-dominated improvement procedures, when considering both solution quality and computation time, were then identified.

BRANCH-AND-BOUND ALGORITHM

This section describes a branch and bound procedure that finds a sequence for a set of jobs that minimizes the sum of weighted squared tardiness when the jobs have distinct release dates. In our branch and bound procedure, a sequence is constructed starting at the beginning and working forward so a node at the p^{th} level in the branch-and-bound tree corresponds to an initial partial sequence (and the associated sub-problem) for the first p jobs in a sequence. Branching adds a job to the end of an initial partial sequence ($p+1^{st}$ position). From a node at level p, up to q (q = n – p) branches may be generated, one for each job not in the initial partial sequence corresponding to the level-p node from which branching occurs. Before branching takes place, dominance conditions can be checked to see if any of the candidates for the $p+1^{st}$ position in a sequence can be eliminated from consideration. Therefore, less than q branches may be generated. To implement this algorithm, a lower bound on the objective value for the completion of an initial partial sequence is needed. We provide dominance conditions that can eliminate nodes and some which are used to help generate lower bounds in the next subsection and then we provide lower bounds in the following subsection.

The branch-and-bound algorithm developed by Schaller and Valente (2012) for the single-machine weighted squared tardiness problem assumed all jobs were available at the beginning of the planning horizon ($r_j = 0$ for j = 1, …, n). Schaller and Valente (2012)'s branch-and-bound algorithm worked from the end of a sequence (a post-partial sequence) to the beginning. Since the jobs did not have release times that would cause idle time on the machine the start and completions times for jobs assigned in a post-partial sequence were known. Since in many cases the jobs assigned at the end of the sequence contribute greatly to the objective (because of high tardiness) this structure has benefits. In the problem considered in this research, since the release times of jobs could cause some idle time on the machine, we would not necessarily know the start and completion times of jobs assigned to a post-partial sequence, and hence we use a different structure in the branch-and-bound tree, in that a node in the tree is represented by an initial partial sequence. In the following subsections, we show how the conditions identified in (Schaller and Valente, 2012) can be used, with some modification. We also point out some conditions that are specific to the problem with release dates that can also be used to truncate the branch-and-bound tree.

Dominance Conditions

In the following, let σ be an initial partial sequence for a node in the branch-and-bound tree and U the set of jobs not included in σ. Let t equal the completion time of the last job in the partial sequence σ and max_r be the maximum release time among the set of jobs to be sequenced ($max_r = max\ \{r_j, j = 1, …, n\}$). Also, let jobs j and k belong to the set U.

(Schaller and Valente, 2012) prove the following for the problem without release dates, if $p_j \le p_k$, $d_j \le d_k$ and $w_j \ge w_k$ then job j is sequenced before job k in at least one optimal sequence. This property does not necessary hold with release dates because there could end up being idle time on the machine between jobs j and k. But if $t \ge max_r$ (the completion time of the last job in the initial partial sequence is greater than or equal to the maximum release time) then there cannot be any additional idle time in the schedule and the property would hold among the unscheduled jobs in the set U, which leads to the following condition.

Condition 1. If $t \ge max_r$, $p_j \le p_k$, $d_j \le d_k$ and $w_j \ge w_k$ then job j is sequenced before job k in at least one optimal sequence.

A second property proved by Schaller and Valente (2012) that can be converted to the problem with release times by adding the criterion $t \ge max_r$ is: If $p_j \le p_k$, $d_k < d_j$, $w_j > w_k$, $(t + p_k) > d_j$ and $w_j * (2*max\ \{t + p_k – d_j, 0\} + 1) \ge w_k * (2*max\ \{t + p_k – d_k, 0\} + 1)$ then job j precedes job k in an optimal sequence. This is condition 2.

Condition 2. If $t \ge max_r$, $p_j \le p_k$, $d_k < d_j$, $w_j > w_k$, $(t + p_k) > d_j$ and $w_j * (2*max\ \{t + p_k – d_j, 0\} + 1) \ge w_k * (2*max\ \{t + p_k – d_k, 0\} + 1)$ then job j precedes job k in an optimal sequence.

A third property proved by Schaller and Valente (2012) that also needs the criterion $t \ge max_r$ in order to convert it to the problem with release times is: If $p_j \le p_k$, $d_j < d_k$, $w_j < w_k$ and $w_j * (2*max\ \{\sum_{l=1}^{n} p_l – d_j, 0\} + 1) \ge w_k * (2*max\ \{\sum_{l=1}^{n} p_l – d_k, 0\} + 1)$ then job j precedes job k in an optimal sequence. Where $\sum_{l=1}^{n} p_l$ is the total processing time of the jobs and represents the maximum completion time for any job when there are no release times. In order to convert this property to the problem with release times, we need an expression that represents the maximum completion time for any job, in addition to the criterion t ≥ max_r. We use maxct to denote an upper bound on the completion time of the last job in a sequence. At the root node of the branch-and-bound tree, the upper bound is maxct = $max_r + \sum_{l=1}^{n} p_l$. As jobs are added to a sequence, maxct is reduced by any processing that occurs before max_r. Condition 3 is the property used in eliminating branches from consideration.

Condition 3. If $t \geq max_r$, $p_j \leq p_k$, $d_j < d_k$, $w_j < w_k$ and w_j * (2*max {maxct – d_j, 0} + 1) ≥ w_k * (2*max {maxct – d_k, 0} + 1) then job j precedes job k in an optimal sequence.

The following conditions determine when one job should precede another when the two jobs are to be sequenced adjacent to each other. Assume there are two j and k that will be sequenced adjacent to each other. Let S be a sequence in which jobs j and k are sequenced adjacent to each other with job j sequenced before job k. Let S' be a sequence that is the same as S except the positions of jobs j and k are exchanged in S'. Let B (jk) be the set of jobs that are sequenced before the jobs j and k in S and S'. Let $C_{B\,(jk)}$ be the completion time of the jobs in B (jk). Let t_s be the start time of job j in sequence S and t_c be the completion time of job k in sequence S. t_s = max {$C_{B\,(jk)}$, r_j}.

Condition 4. $r_j \leq r_k$ or $r_j \leq C_{B\,(jk)}$, $d_k \geq t_c$, and $d_j < t_c$ then job j precedes job k.

Condition 5. $r_j \leq r_k$ or $r_j \leq C_{B\,(jk)}$, $d_k \geq t_c$, and $d_j < d_k$ then job j precedes job k.

Condition 6. $r_j \leq r_k$ or $r_j \leq C_{B\,(jk)}$, $d_k \geq t_c$, and $r_k > t_s$ then job j precedes job k.

Let $t_{s'}$ be the start time of job k in sequence S' ($t_{s'}$ = max {$C_{B\,(jk)}$, r_k}. If d_j and $d_k < t_c$, we have the following condition.

Condition 7. $r_j \leq r_k$ or $r_j \leq C_{B\,(jk)}$, $d_j < t_c$, $d_k < t_c$, and (w_j * max {$t_s + p_j - d_j$, 0}^{2} + w_k * max {max {$t_s + p_j$, r_k}$_+$ p_k + – d_k, 0}^{2}) < (w_j * max {max {$t_{s'} + p_k + p_j - d_j$, 0}^{2} + w_k * max {$t_{s'} + p_k - d_k$, 0}^{2}), then job j precedes job k.

Let k be a job in an initial partial sequence σ in which there is idle time immediately after job k and before the next job sequenced in σ. Let it = the idle time immediately after job k. Let C_b = the completion time of the job sequenced immediately before job k (if job k is the first job sequenced in σ, then C_b = 0). If there is an unscheduled job j (j ϵ U) such that $p_j > p_k$, either $r_j \leq r_k$ or $r_j \leq C_b$, and (max {r_j, C_b} + $p_j - C_k$) ≤ it, the following three conditions hold.

Condition 8. If $d_j \leq d_k$ and $w_j \geq w_k$ then the partial sequence σ can be eliminated from further consideration.

Condition 9. If $d_j < d_k$ and $w_j < w_k$, and w_j * (2*max {maxct – d_j, 0} + 1) ≥ w_k * (2*max {maxct – d_k, 0} + 1) then the partial sequence σ can be eliminated from further consideration.

Condition 10. If $d_j > d_k$ and $w_j > w_k$, $d_j < C_k$, and w_j * (2*($C_k - d_j$) + 1) ≥ w_k * (2*($C_k - d_k$) + 1) then the partial sequence σ can be eliminated from further consideration.

Let t equal the completion time of the last job in a partial sequence σ and let jobs j and k belong to the set U.

Condition 11. If $p_j > p_k$, $d_j \leq d_k$, $w_j \geq w_k$, $r_j < r_k$, $r_j \geq t$, and $((r_k - r_j) + p_k) > p_j$ then job j precedes job k in an optimal sequence.

In addition, the set of active schedules will contain an optimal schedule. Let σ be an initial partial sequence and t equal the completion time of the last job in the partial sequence σ. Let minct = the minimum completion time among the jobs in the set U if they were to be scheduled next, immediately after the initial partial sequence σ. The following condition holds.

Condition 12. Let k be a job in the set U. If r_k > minct then the initial partial sequence σk can be eliminated from further consideration.

Lower Bounds

When branching occurs and new nodes are created, a lower bound on the sum of weighted squared tardiness that would be obtained by the completion of the initial partial sequence corresponding to those nodes is calculated. If the lower bound is less than the lowest sum of weighted squared tardiness found so far for complete sequences (incumbent value) and the node does not represent a complete sequence, the node is retained for additional branching. If the lower bound is less than the incumbent value and all the jobs have been sequenced in the branch ending with the node (the node represents a complete sequence), then the incumbent value is updated to equal the lower bound, the sequence is recorded and the node is eliminated. If the lower bound is greater than the incumbent value, the node is eliminated. The algorithm uses a depth first strategy. This strategy selects for branching the node at the lowest level of the tree, breaking ties by choosing the node with the lowest lower bound.

A lower bound can be obtained by just considering the jobs in σ. Let Z (σ) equal the sum of the weighted squared tardiness of the jobs in σ. The total weighted squared tardiness of a complete sequence which has σ as an initial partial sequence can be no lower than Z (σ), therefore Z (σ) is a lower bound. This lower bound is strengthened by including a job based lower bound for the jobs in set U. A lower bound for the completion time for each job j in the set U is max $\{t, r_j\} + p_j$. A lower bound on the weighted squared tardiness of the jobs in the set U is $\sum_{j\in U} w_j * \max\{(\max\{t, r_j\} + p_j) - d_j, 0\}^2$. This lower bound can be further improved if we can identify jobs that belong to the set U that will precede job j in an optimal sequence. Define B(j) as the set of jobs known to be sequenced before job j in at least one optimal sequence. Let $t_{B(j)}$ be the completion time of the jobs in the set B(j). Max $\{t_{B(j)}, r_j\} + p_j$ is a lower bound on the completion time of job j, and max $\{(\max\{t_{B(j)}, r_j\} + p_j) - d_j, 0\}^2 * w_j$ is a lower bound on job j's weighted squared tardiness in an optimal sequence To obtain the lower bound on the completion of an initial-partial sequence σ (LB_σ), we add the job based lower bounds for the jobs in set U to Z (σ):

$$LB_\sigma = \sum_{j\in U} \max\{(\max\{t_{B(j)}, r_j\} + p_j) - d_j, 0\}^2 * w_j + Z(\sigma).$$

COMPUTATIONAL RESULTS

We performed computational tests of the algorithm to see the types of problems it would be practical to utilize the algorithm and also how the practicality of using the branch-and-bound algorithm for this problem compares to that of the problem without release dates.

Instances

The computational tests were performed on randomly generated instances. The process used to generate the instances is in accordance with previous works concerning single machine tardiness problems with release dates.

The instances were generated as follows. We considered four levels of number of jobs (n): 10, 15, 20, and 25. For each job j, an integer processing time p_j was generated from a uniform distribution $[1{,}100]$, and an integer weight w_j was generated from a uniform distribution $[1{,}10]$. In what regards the release dates r_j, for each job j an integer release date was generated from the uniform distribution $\left[0, \alpha \sum_{j=1}^{n} p_j\right]$, where α is the release date range parameter. The following four values were considered for α: 0.25, 0.5, 1.0, 1.5.

The due dates d_j are not generated directly. Instead, for each job j we generate an integer slack time between the due date and the earliest possible completion time, that is, we generate an integer value for $d_j - (r_j + p_j)$. The due date is then obtained by adding the slack value to the earliest possible completion time. The integer slack time is generated from the uniform distribution $\left[0, \beta \sum_{j=1}^{n} p_j\right]$, where β is the due date range parameter. The following three values were considered for β: 0.05, 0.25, 0.5.

For each instance size n, we then have 12 combinations of release date range α and due date range β, thus providing a wide variety of conditions. In the full set, 10 instances were generated for each combination of instance size n, release date range α and due date range β, therefore giving a total of 120 instances for each problem size (number of jobs, n).

Results

The algorithm was performed for each instance. A 300 second time limit was used for each instance. The number of seconds used to find an verify an optimal solution was recorded for each instance. If the algorithm terminated before completion, the number of seconds was recorded as 300. In the table that follows, the average seconds per instance and the number of instances solved within the time limit for each set by the number of jobs (n) is shown.

Table.

	Number of Jobs							
	10		15		20		25	
Set	Ave.	# Solved	Ave.	# Solved	Ave.	# Solved	Ave.	# Solved
1	0.07	10	1.51	10	191.26	6	295.41	1
2	0.05	10	1.54	10	79.77	10	300.00	0
3	0.04	10	1.06	10	100.82	9	281.67	1
4	0.07	10	3.45	10	90.12	9	300.00	0
5	0.06	10	1.59	10	229.32	4	254.04	2
6	0.06	10	1.16	10	192.79	8	286.88	1

7	0.04	10	0.19	10	9.68	10	4.38	10
8	0.02	10	0.26	10	1.91	10	7.84	10
9	0.03	10	0.05	10	4.34	10	5.42	10
10	0.02	10	0.03	10	1.95	10	0.10	10
11	0.03	10	0.03	10	0.64	10	0.07	10
12	0.02	10	0.04	10	0.35	10	0.08	10

The results show that, as expected, as the number of jobs (n) increases the algorithm requires more time to find and verify an optimal solution. Also, as the number of jobs increases, a greater number of instances fail to complete within the 300 second time limit. The algorithm was able to solve all of the instances with 10 or 15 jobs, 106 of 120 instances with 20 jobs, and 65 of 120 instances with 25 jobs.

These results would indicate that this problem may be more difficult to solve in a reasonable amount of time compared to the problem without release dates. Schaller and Valente (2012)'s algorithm was able to solve all instances up to 40 jobs and almost all instances with 50 jobs.

The release date range parameter (α) also affects the performance of the algorithm. As α increases, the algorithm needs less time to solve instances. When $\alpha > 0.5$, the algorithm averages less than 10 seconds per instance for all problem sets. When $\alpha = 1.5$, the algorithm averages less than 2 seconds for all problem sets. The algorithm is able to solve all the instances when $\alpha > 0.5$. When $\alpha \leq 0.5$ and $n = 25$, the algorithm only solves 5 of the 60 instances within the 300 second time limit. The reason the algorithm is able to solve more instances and require less time when α increases is that the algorithm checks that only active schedules are considered by using a dominance condition.

CONCLUSION AND FUTURE RESEARCH

This paper addresses the problem of minimizing total weighted squared tardiness on a single-machine when jobs have release dates. A branch-and-bound algorithm is proposed for small sized instances. This algorithm can be used to obtain optimal solutions whose objective values can be compared to the objective values generated by heuristics in order to gage the accuracy of the heuristics. Several conditions are provided that can help truncate the branch-and-bound tree. Also, a lower bound is proposed to help truncate branches in the tree. The algorithm is shown to be able to solve instances with 10 and 15 jobs but becomes time consuming for several sets of instances when the number of jobs is greater 15.

For future research, the development of additional conditions or ways to improve the lower bound can be explored. Also, alternative structures of the branch-and-bound tree can be investigated. For the problem without release dates the tree worked from the end of a sequence to the beginning. In our algorithm we work from the beginning because if we started at the end, we would not know the completion times of the jobs. This can be explored. We can develop lower and upper bounds on the completion times and these can be used to develop job based lower bounds.

REFERENCES

Costa, M.R.C., J.M.S. Valente, and J.E. Schaller, Efficient procedures for the weighted squared tardiness permutation flowshop scheduling problem," *Flexible Services and Manufacturing Journal*. (July 2019): 1 – 36.

Fisher ML. Optimal solution of scheduling problems using Lagrangian multipliers: part I. Operations Research; 1973a; 21 (5); 1114 – 1127.

Fisher ML. Optimal solution of scheduling problems using Lagrangian multipliers: part II. Operations Research; 1973b; 21 (6); 294 – 317.

Gonçalves T.C., J.M.S. Valente, and J.E. Schaller, "Metaheuristics for the Single Machine Weighted Quadratic Tardiness Scheduling Problem," *Computers and Operations Research*, 2016.

Hoitomt DJ, Luh PB, Max E, Pattipati KR. Scheduling jobs with simple precedence constraints on parallel machines. IEEE Control Systems Magazine 1990; February; 34 – 40.
Schaller, J. E., and J.M.S. Valente, "Minimizing the Weighted Sum of Squared Tardiness on a Single Machine," *Computers and Operations Research*, vol. 39, no. 5, p. 919, 2012.

Schaller, J.E. and J.M.S Valente, "Efficient heuristics for minimizing weighted sum of squared tardiness on identical parallel machines," *Computers and Industrial Engineering*, vol. 119, pp. 146 – 156, 2018

Sun X, Noble JS. An approach to job shop scheduling with sequence-dependent setups. Journal of Manufacturing Systems 1999; 18; 416 – 430.

Sun X, Noble JS, Klein CM. Single-machine scheduling with sequence dependent setup to minimize total weighted squared tardiness. IIE Transactions 1999; 31; 113-124.

Taguchi G. Introduction to quality engineering. Asian Productivity Organization: Tokyo, Japan; 1986.

Thomalla CS. Job shop scheduling with alternative process plans. International Journal of Production Economics 2001; 74; 125–134.

Valente, J.M. and J.E. Schaller, "Dispatching Heuristics for the Single Machine Weighted Quadratic Tardiness Scheduling Problem," *Computers and Operations Research* vol. 39, no. 9, p. 2223, 2012.

Wagner BJ, Davis DJ, Kher H. The production of several items in a single facility with linearly changing demand rates. Decision Sciences 2002; 33; 317–346

DECISION SCIENCES INSTITUTE
Stringdown: an online exercise to introduce scrum

Brad C. Meyer
Drake University
2507 University Avenue, Des Moines, IA 50311
bradley.meyer@drake.edu
(515)-271-2490

ABSTRACT

In the midst of the pandemic, confined to an online environment, we used Google Jamboard to implement a team-based activity to demonstrate many of the salient aspects of the agile approach to project management. Students worked with stories and story value, sprints, estimation of points for a sprint, retrospectives, and continuous improvement. We describe the exercise and compare to the in-person Lego Scrum exercise we were using before COVID.

KEYWORDS: Active learning, Innovative education, Scrum, Agile

INTRODUCTION

One of the critical topics in the core operations management course of most business schools is project management. Especially In the non-manufacturing sector, projects are an essential avenue to accomplishing the organization's work. CPM, PERT and project crashing are still useful topics, but we would be remiss to leave out coverage of agile and scrum. Even though these methodologies cut their teeth in software development, their use continues to spread to other work types. We added the topic to our OM class several years ago when an industry trainer offered to run a scrum exercise in our class.

This exercise required groups to build an airplane with Legos according to a set of specifications from a customer(Paz, 2018). The requirements were described in stories with business value points, and several of the requirements changed over the course of the 4 sprints. The requirements were provided to the groups in a simple statement they could write on a sticky note to put on their scrum board. The statements appeared to be complete, but more specificity could be uncovered if you raised questions to the customer. This active learning exercise was well received by the students and brought understanding of the concepts of customers, product owners, sprints, stories, business value, sprint planning, retrospective, continuous improvement, scrum boards, prioritization of work, backlogs, and the dynamics of project requirements. The briefing also described waterfall vs agile, and the value of teams.

After using the exercise several times and learning how to run it ourselves, we could see several shortcomings. First, the exercise was long. We would allocate an entire 75 minute class period to the exercise and had to brief it in a previous period. Second, the activity seemed to have one main message: talk to your customer! The scoring of teams at the end of the exercise was based on criteria known to the customer and only known to the teams if they asked enough questions. The winning teams were those who learned the most details from individual questions they asked the customer. This is a great lesson and a major strength of the agile approach to projects. But the complexity of the task (building an airplane with lots of requirements) in the context of 4 short cycles of sprint/retrospective did not incline the groups to examine their work methods. Their creative energy was expended on the product, not on the process.

In the spring of 2020, we dropped the exercise because it was scheduled shortly after our in-person classes were shut down due to the pandemic and there was not enough time to develop or find a suitable online replacement. During the fall, we began to search for an online scrum activity. The option that had the most potential was an adaption of the ball scrum activity(May et al., 2016). In the ball activity, groups of 6 to 12 people pass balls through the hands of all members according to a few restrictive rules. The final score is the number of balls they can pass in the time of a sprint. Someone had adapted this game to create an online exercise using a Google document(Florit, 2020). The Google document was editable by all the group members and instead of balls, text strings were passed from one grid square to another, where the grid was a table in the Google document. Each person was assigned one grid square and when the text string appeared in their grid square they would edit it by adding their number to the end of the string and then they would cut and paste the string to another person's square. They incorporated similar rules to the ball pass game by making it illegal to put the string in the grid box immediately to the left or right of the player's grid box.

The task in this game is simple enough that the players can devote energy to improving the process. However, every string is the same, so there is no implementation of business value, and no prioritization of stories. There is no scrum board, but the groups do perform the task of estimating the number of strings they will process in the upcoming sprint. We started from the structure of this game and enhanced the exercise in several ways, over a few iterations. We first tried using the collaborative whiteboard in Blackboard Collaborate Ultra but moved to Google Jamboard. In our first iterations, the players would build text strings by adding letters. To implement priorities, we changed to pre-specified text strings of different lengths that would be "processed" by changing upper case letters to lower case letters. The exercise, in its present form, is described below after a brief discussion of relevant literature.

LITERATURE REVIEW

According to Rigby, Sutherland and Takeuchi(Rigby et al., 2016), scrum arose when Jeff Sutherland was charged with the audacious task of developing a new major software product for the Easel Corporation in less than six months. His research into an appropriate development methodology led him to an article by Takeuchi and Nonaka (Takeuchi & Nonaka, 1986) describing how Japanese manufacturing companies were greatly reducing product development time using ideas that arose from what was later termed "lean methodologies." In particular, Sutherland was intrigued by the analogy of a group performing like a rugby team, instead of a relay team. In a relay team, each person performs separately, and each runner passes a baton once to another runner. In a rugby team, the team performs collectively, exchanging the ball repeatedly as they move towards the goal.

Inspired by these ideas, Sutherland developed a software development strategy that was successful for the Easel project, and then for several companies afterwards(Jeff Sutherland & Sutherland, 2014). In 1995, his methodology, which he called scrum in accordance with the rugby analogy, was presented at the annual OOPSLA (Object-Oriented Programming, Systems Languages and Applications) conference, with colleague Ken Schwaber(Jeffrey V Sutherland et al., 2012). A few years later, a group of software developers created the Manifesto for Agile Software development(Beck et al., 2001). The agile perspective or philosophy embodies 4 basic tenants:

- individuals and interactions over processes and tools
- working software over comprehensive documentation
- customer collaboration over contract negotiation
- responding to change over following a plan.

Scrum is a recognized methodology within agile. There is extensive literature in the software development and system design field on agile and scrum. The interested reader is referred to (Dikert et al., 2016; Cardozo et al., 2010).

As scrum has become more popular in practice, educators who teach software development have begun to create active learning exercises so students can learn concepts of scrum through experience. The ball game (May et al., 2016) and the Lego airplane game (Paz, 2018) were described above. There are also other Lego exercises(Hof et al., 2017; Paasivaara et al., 2014; Steghöfer et al., 2017), an origami exercise(Sibona et al., 2018), and a card game(Fernandes & Sousa, 2010) described in the literature. (Naik et al., 2020) describe using an open-source groupware tool, Trello, to support agile scrum and several authors describe running the class project with a scrum methodology. See (May et al., 2016) for a listing of articles on using scrum for managing class group projects.

Outcomes reported from scrum activities in class include 1) more motivation to learn about and apply scrum, 2) common experience to refer to in future classes where scrum ideas are covered, 3) increased interest from recruiters, 4) improved groupwork skills, and 5) improved class project outcomes.

GAME DESCRIPTION

The Stringdown scrum activity requires groups of students to process stories which consist of text strings of capital letters. Each group member is assigned one letter and processes the text string by changing all of the uppercase of that letter in the string to lower case. In each sprint, the group is presented with a new set of text strings, each of which has an assigned business value. The group members are together in a Zoom breakout room (or a comparable online collaboration space) and do their processing in a set of frames in Google Jamboard. See Figure 1.

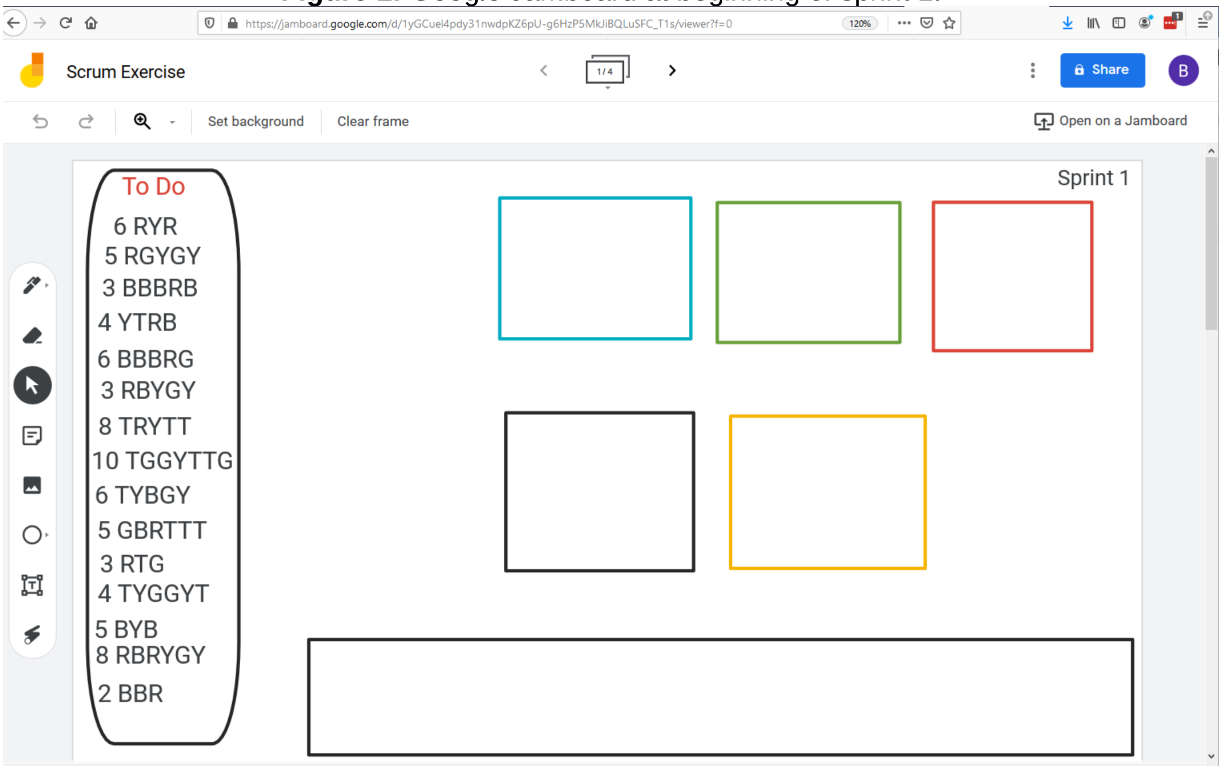
One of five colored boxes is assigned to each of 5 group members and the letter assigned to each group member is the first letter of the color of the box, (T)eal, (G)reen, (R)ed, (B)lack, and (Y)ellow. The text strings can be dragged from location to location and double-clicking on a text strings makes it editable. Once all the letters in the text string are changed to lowercase, the string can be dragged to the large done box at the bottom of the frame.

The business value of each story is the number proceeding the letters in the string. In general, the more letters, the higher the value, but there are exceptions to this which suggests that a prioritization scheme might allow more points in less time.

The groups perform the processing task during 4 sprints. Each sprint is preceded by a retrospective and planning meeting during which the team is encouraged to sum the business points earned in the previous sprint, consider the process, improve it, estimate the number of business points they will achieve in the upcoming sprint, and make any prioritization decisions regarding the stories in the upcoming sprint.

For each sprint, the retrospective/planning meeting is timed at 2 minutes and the processing is given 1.5 minutes. In moving from sprint to sprint, the group moves to the next frame in the Jamboard.

Figure 1: Google Jamboard at beginning of sprint 1.



Each group must also designate a timekeeper, an inspector (who ensures that all the done strings have no uppercase letters), and a recorder, who records the business value estimates and completed score for each sprint. These roles could overlap with the processor roles described above, allowing teams of 5 to 8 players. It is also possible to assign one team member the role of selecting items from the to do list and dragging them to the first processing box.

Before the game, students need a briefing session where they can be oriented to Google Jamboard, the task to be done, the roles, the retrospective and planning process, and the time flow requirements of the activity. Appendix 1 provide the current handout that students read before class. We roughly walk through the handout for the briefing. This briefing is in addition to an introduction to agile and scrum presented in class. At the current date, our operations management text only mentions scrum, giving the impression that scrum is simply a daily meeting to discuss progress on projects.

To date, we have not developed an extensive debrief to the exercise, but we ask the students to describe what they learned from the activity. Example responses are provided in the next section.

GAME EXPERIENCE

Figures 2, 3, and 4 show example screens at the end of a sprint. They show some of the improvement steps taken by the groups. In Figure 2, the group was not able to process all the strings and left three in the green box and one in the To Do box.. There is a Y in the yellow box because that apparently was a hard letter to remember. In Figure 3, the group took a more drastic approach to providing the letters to the players. In Figure 4, the team decided that the boxes were a hindrance and removed them all together, replacing the box with a heading name of the color. The 83/128 was added by the team at the end of the sprint, showing points earned compared to points possible.

Figure 2: Example end of sprint Jamboard screen

Figure 3: Example end of sprint Jamboard screen showing improved letters

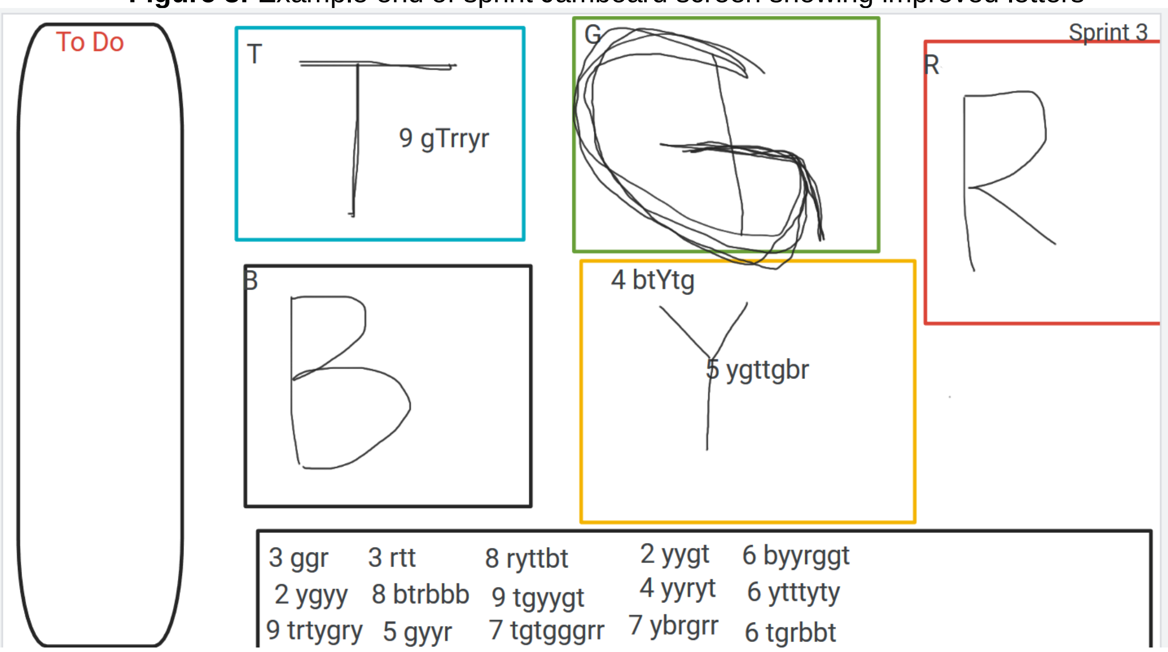
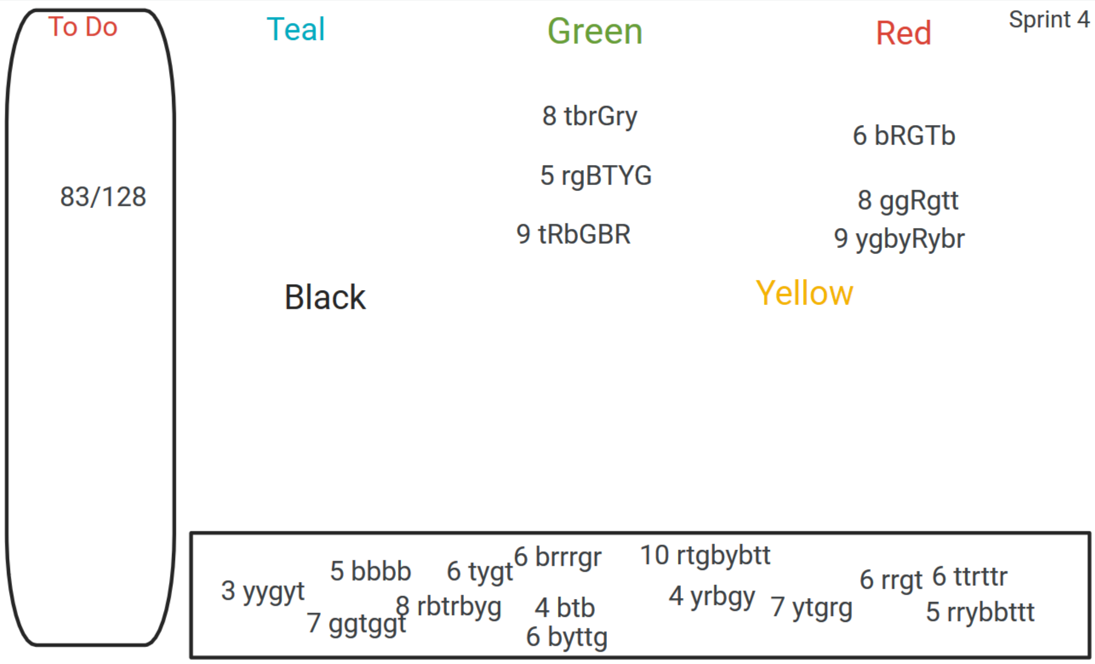


Figure 4: Example end of sprint Jamboard screen showing improvement of removing boxes



The student reaction has improved with each iteration of the game. It was especially good in the most recent version when the time was short enough that the groups could not get all of the strings processed. That increased the drive for improvement, and consequently, increased the collaboration. Here is one's student's reflection on the assignment that brings out the story of continual improvement.

> "I learned a lot in this group activity. My team adopted at least one strategy after every sprint in order to improve our numbers/results. We found that we piled up the letters in certain boxes and the people in charge of those boxes would get overwhelmed. We decided to make sure this didn't happen again by putting the letters in a different box if one was already full. We also decided to increase the size of the text boxes to make more room for numbers. The strategy that gave the best results was picking letters that had the biggest number values before the letters with the smaller values to increase our score. It was a very fun experience that taught me a lot about teamwork and how to cooperate to better the team's agenda."

When the activity was introduced, I mentioned the concept of collaboration, but did not emphasize it, so it speaks to the value of the assignment to find that most of the student reflections included statements on collaboration and teamwork. For example,

> "I learned how impactful collaboration and communication is in group work, specifically for projects…it encourages all thoughts and ideas from all members of a team and thus increases ideas and exploration for greater productivity."
>
> "The most effective way was the final round when we on the speaker to communicate while shifting everything around. I think that all these strategies will make sense if we approach to running projects in the future, as the most efficient way is to communicate with each other, discussing about the problems and solutions together."
>
> "It was also important to listen to everyone in between tasks so that we could hear about other people's struggles and what they did to solve them."

> "Those of us who communicated our needs out loud rather than trying to solve it on our own were more efficient as a collective than we were individually."

There were numerous comments about continuous improvement,

> "What I learned from the scrum activity today is that you can always improve your process."
> "We were able to improve our score each time."

There was at least one student, who came up with a misperception and wrongly attributed to the methodology a problem that the group could have solved.

> "I did notice that Red and Teal had more work to do than me (yellow). I often didn't have anything to do and had to watch my other teammates struggle to keep up with how many strings they had to work on. I think that if there is a way to handle this idle time and sudden increase of work for others, this way of running a project could be really useful."

This is a danger of using active learning exercises where students are to deduce lessons from the experience. But to be fair, misunderstandings also arise from lecture material. Also, this is why we grade assignments and provide feedback to the students.

As for improvements that teams described in their write ups, the list included

- talked to each other instead of performing our jobs in isolation
- improved the process of placing strings to be processed by making the boxes bigger, displaying the letter associated with the box, remove the border on the box, and creating rules to eliminate congestion
- prioritized the strings to improve points earned; one group moved a prioritization task into the retrospective/planning time, physically rearranging the list of strings to put the strings in priority order from top to bottom.
- one group categorized the strings in the to do list by first letter, to improve the movement from the list into the processing area.

Some students commented on the value of scrum.

> "I do believe that this activity showed me that scrum can be a useful and effective strategy."
> "The scrum approach to running projects could be extremely useful in today's business environment, especially in an office setting. Many offices do not have collaborative work areas and could greatly benefit from the accountability that scrum provides."
> "I found that the timed work period *(i.e., the set length of the sprint)* forced me to be more focused than I would have been otherwise."

Finally, the activity did have some aspect of the motivation of enjoyment associated with it.

> "This activity honestly ended up being a lot of fun!"
> "Thank you so much professor for such a fun activity, it really relaxes my head a lot with so much pressure going on right now." *(A COVID-19 lockdown sentiment.)*

We have not yet developed exam questions to test the learning from this activity, but that will be necessary in order to evaluate the efficacy of the exercise.

CRITIQUE AND FUTURE IMPROVEMENTS

The strengths of this game are the following:

- it is a group activity that can be used in a synchronous online class to give an experience with scrum.

- the activity itself, including opening instructions, can be done in 30 minutes of class time.
- the activity demonstrates sprint cycles, stories, a meager scrum board, need for collaboration, continuous improvement, business value, prioritization of stories, retrospective, sprint planning
- the activity can be done at no cost, given Google accounts and an online collaboration tool such as Zoom.

There are some components of scrum that are not demonstrated by this game. First, the activity does not have a customer, so the critical lesson of continual interaction with the customer is not taught by this exercise. Second, the work done in the task is not of the project nature; it is rather highly repetitive, with the tasks being set and fixed. There is nothing creative about the product and there is no final coming together of all the separate tasks into a final major product. The stories are not really stories in the sense of little plot lines that get built into the final result.

Interestingly enough, the missing aspects of this exercise are the strengths of the in-person Lego airplane activity we did before the pandemic. The interaction with the customer, for example, was a strength of the in-person game. Moving customer interaction to an online game has challenges. In person, you can have one customer for multiple groups since access is free and with no lag time. With groups into Zoom rooms, the sharing of a customer is cumbersome. There is a time lag to enter or leave a room that would make multi-group discussions with a customer infeasible given a 2 minute retrospective/ planning meeting. But it might be possible to use a text messaging system to communicate with the customer. Of course, the Stringdown game would have to be redesigned to make customer questions relevant to the work.

In order to make the task list more project like, a tool like Minecraft could be utilized. The students could construct virtual objects that worked together. The drawback would be the larger learning curve for a tool of this type and the loss of enough process consistency to spur ideas for process improvement.

One minor problem with Stringdown is that there is no test to make sure that the final strings match the original string, but with lowercase letters. Thus, if the editing process removes or adds a letter, or changes the wrong letter, there is no way to know that. If such an inspection could be implemented, this could be compared to finding software bugs and would add an element of customer satisfaction. We have done no tracking to see the degree to which these kinds of errors occur.

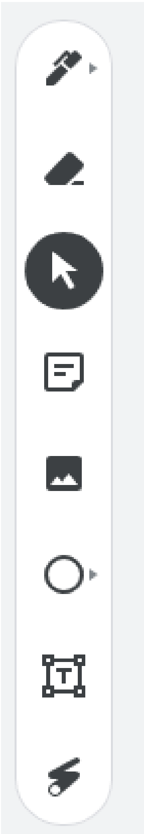
CONCLUSION

We developed an online activity to give students an experience with the scrum methodology. It develops an understanding of many, but not all, important concepts in scrum. The activity is short enough to be used without displacing a large amount of content. Students received the activity well and commonly stated that the exercise emphasized to them the power of collaboration and the value of continual process improvement. This exercise is relevant to operations management and is a good example of how concepts we commonly teach in Lean can be applied to project management.

APPENDIX 1: Class handout to prep for the class period where the exercise is conducted

Prep for the activity

1. Understand the basics of the Google Jamboard app. (Google account required to use Jamboard)



There is a pen tool, an eraser tool, a select tool, a sticky note tool, an image insert tool. a shapes tool, a textbox tool, and a laser tool. You will be mainly using the Select tool for this activity to select text strings, drag them, and change them.

If your team finds it useful to add anything to the screen, you could use some of the other tools. They behave very much like drawing tools in other software.

One important thing to note if you add anything to the screen is that some shapes, sticky notes, or textboxes you add will sit on top of other things and hide them. If that creates a problem, try setting the fill to transparent or clicking on the item menu, choosing Order and Send to back. That makes everything else sit on top of the item.

If you accidentally create an item, just select it and push your keyboard's delete key.

There is a scaling (zoom in/out) control at the top of the canvas. This can be used if you need to change the size of the page for your screen.

2. Select three positions for your team: the inspector, the time keeper, the recorder.

The inspector will make sure that all the other players have processed the text strings that are in the Done box.

The time keeper needs to have a handy timer, like on a smart phone, to time each iteration at 1.5 minutes of processing time and each planning/retrospective session at 2 minutes.

The recorder makes a simple table with Sprint number 1 through 4 and then records the estimated number of business value points that will be processed in the next sprint and after the sprint, the actual number of business value points that were achieved.

3. You will be moved into breakout rooms so you can discuss and work together. Five team members will each choose a color, which will have a corresponding letter and box. The inspector should not also have a color, but if there are less than 8 players in your team, the recorder, and the timer could also have a color.

Black (B) Red (R) Yellow (Y) Green (G) Teal or Turquoise (T)

The whiteboard begins with a To Do box full of "stories," five colored processing boxes, and a Done box. The To Do box should be left as is, but the team may choose to resize or move the other boxes.

The team members should work with each other to arrange the workspace in a way that facilitates the process. You may want to wait until after the first round to learn what changes will be helpful. The positions can be changed between sprints or even during sprints to improve performance.

The preparation is now complete.

Running the 4 Sprints

The activity itself will consist of 4 sprints. For each sprint you will have a 2 minute planning/retrospective period and then a 90 second processing period. Each sprint will take place on a different frame in the Jamboard. You move to the new frame during the 2 minute planning/retrospective period.

In the planning period figure out how to best accomplish the task to get as many business value points as possible in the 1.5 minutes of processing. Figure out what didn't work or what could be improved in the next round.

Before the processing round starts, record an estimate of how many points you expect to complete during the upcoming processing period.

The process period proceeds as follows:

In the processing period, items will be moved from the To Do box to one of the colored boxes. The person associated with that color will change any of their letter in the text string from a capital to a lower case. Then the item will be moved to another box whose capital is still in the string, and again, any uppercase letters are changed to lower case, keeping the letters in the same sequence. To change a letter, use the select tool, double click on the string and then you can edit. Click off the string to be able to move it again. Once all the capitals are changed to lower case, the string is dragged to the done box. The inspector will keep an eye on the done strings to make sure they are really done. The team will get no points for a string with a capital letter in it.

Once the time keeper announces that the 1.5 minutes are up, the recorder should add up the points of the completed strings in the done box and record the number. At that point the group can discuss what went well or poorly during the sprint and think of ways to improve for next time.

After the Sprints

The instructor will be checking the breakout rooms and when all groups have finished, will pull everyone back into the main room for a closing discussion.

Appendix 2: Instructor Setup for Scrum on Google Jamboard Exercise

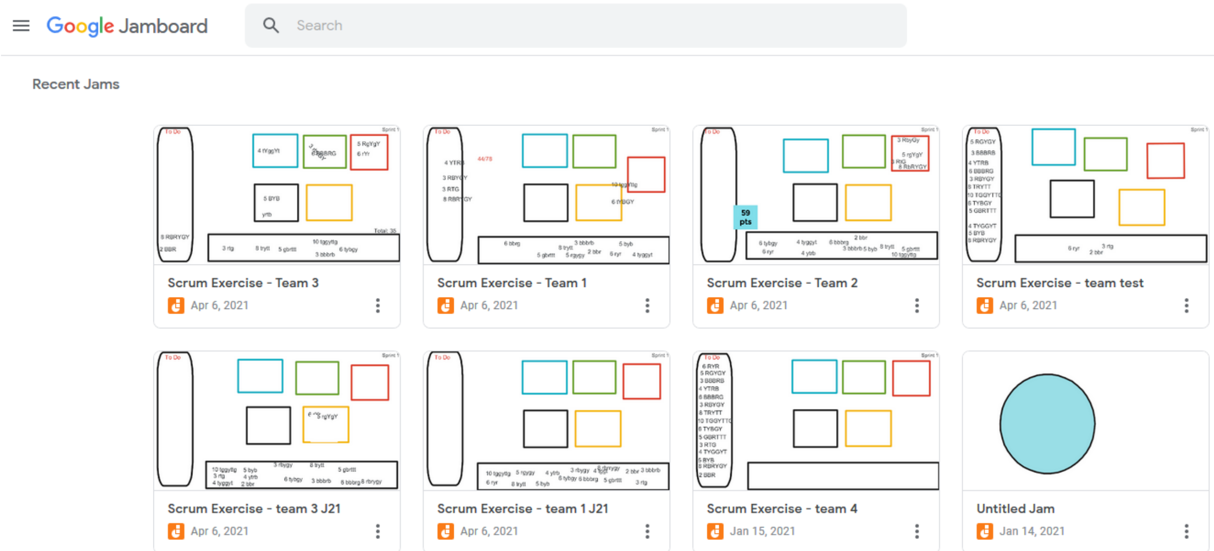
0. Get a copy of the template by following this link:

https://jamboard.google.com/d/1GRdW2BJV5nbeixO2k0U_HXbtcZP0y4wYbCapex521GM/edit?usp=sharing

Once you are at this link, click the 3 dots in the upper right, next to the word Share and choose Make a copy. This will save it to your own Jamboard file area. Next you can set up the groups for your class.

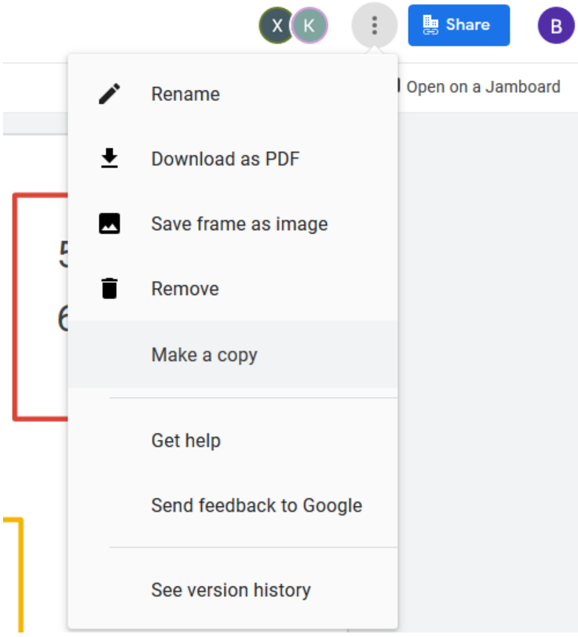
1. First copy the template to other teams

Get to Jamboard and click on the little J symbol (upper left hand corner) to see all your Jamboards.



Click on the “Stringdown Scrum Exercise template” Jamboard (or whatever you named it in step 0 above.)

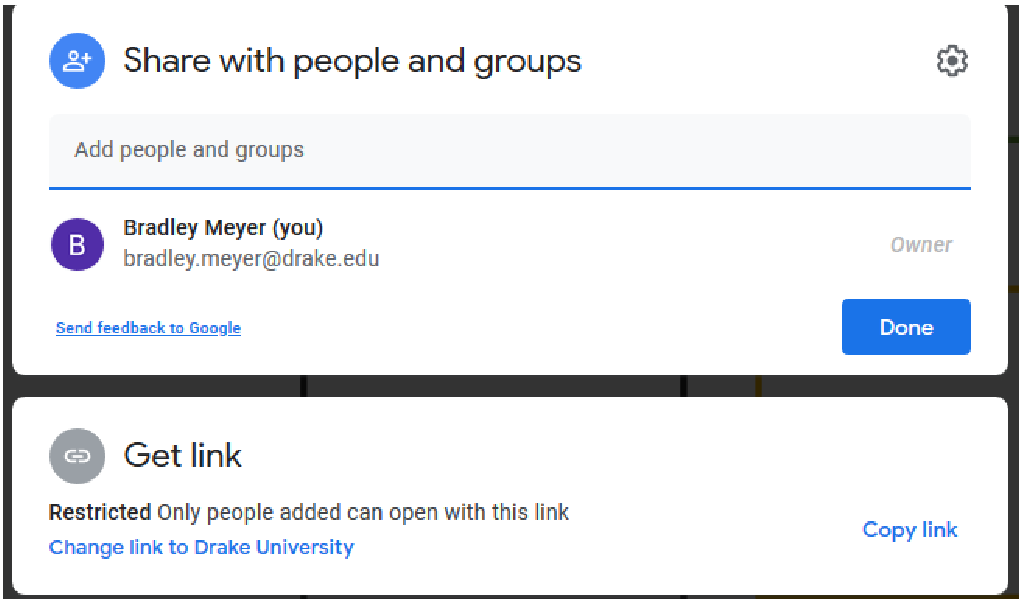
Make a copy for each team. (Three dots, next to the word Share.) It will ask you to name the copy, so it is best to develop some naming convention with team and number.



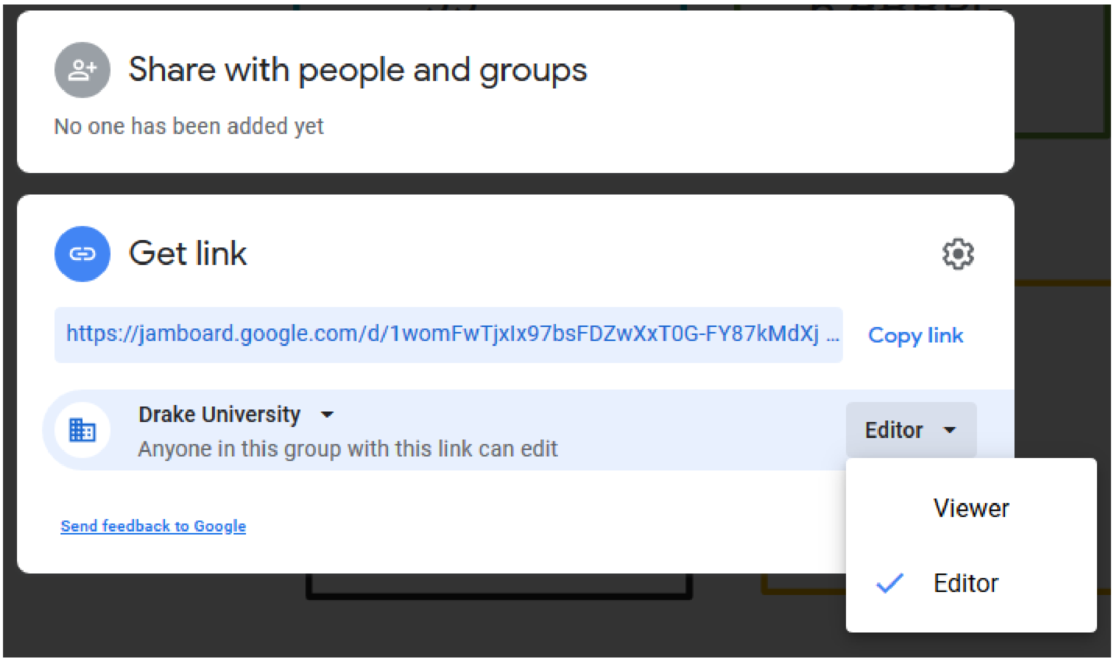
2. For each copy, you have to change the share settings. Set the access to anyone with link and give editor privilege. This requires the following steps:

a. In upper right, click on Share. That pops up the dialog shown below.

b. It's usually too onerous to enter individual names, so go to the Get link section and change link to your university/organization or to anyone with the link.

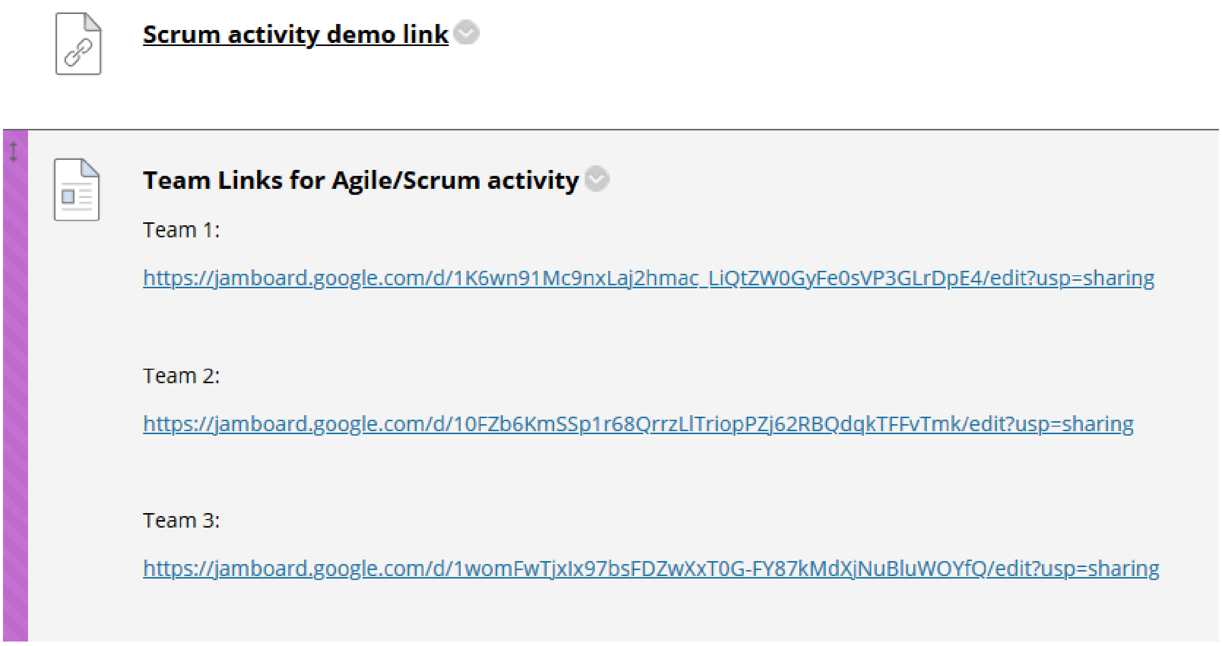


c. To make sure anyone with the link can edit it, find the menu to the right of your university name and change from Viewer to Editor.



Also, at this point copy the link. Then post it for students to use. I use one link for everyone to view for my demo, and then a separate Jamboard and link for each team. By the way, each Jamboard has 4 frames, one for each sprint.

Here is what it looks like on my Blackboard site:



APPENDIX 3: Spreadsheet for creating strings

This is only relevant if you want to create more sprints or change the strings that are in the template. I could not find a slick way to move from Excel to Jamboard, so I had to create text boxes in Jamboard and type or paste the value from each string into a string in Jamboard. For the Jamboard frames, I used a longer set of strings each time, since the groups will be improving as time goes by. Below in yellow are the strings used for the game. I also parsed out any strings of all the same letter.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1										Letter							
2	Color table			Story	Length	1	2	3	4	5	6	7	8	Total	string	points	Coded
3	0	B		1	4	Y	R	R	Y	T	G	Y	T	YRRYTGYT	YRRY	6	6 YRRY
4	0.2	G		2	3	G	G	T	Y	G	B	G	B	GGTYGBGB	GGT	1	1 GGT
5	0.4	T		3	4	Y	R	B	G	T	Y	Y	R	YRBGTYYR	YRBG	3	3 YRBG
6	0.6	R		4	6	G	G	G	T	Y	G	B	R	GGGTYGBR	GGGTYG	9	9 GGGTYG
7	0.8	Y		5	7	R	G	G	Y	T	R	G	Y	RGGYTRGY	RGGYTRG	7	7 RGGYTRG
8	1			6	6	B	B	T	B	T	R	G	G	BBTBTRGG	BBTBTR	7	7 BBTBTR
9				7	5	G	G	G	B	R	B	Y	R	GGGBRBYR	GGGBR	7	7 GGGBR
10				8	6	Y	T	Y	Y	T	B	G	Y	YTYYTBGY	YTYYTB	5	5 YTYYTB
11				9	5	Y	R	R	Y	R	Y	Y	Y	YRRYRYYY	YRRYR	8	8 YRRYR
12				10	4	Y	Y	Y	T	G	T	G	B	YYYTGTGB	YYYT	4	4 YYYT
13				11	7	T	T	G	B	T	T	T	G	TTGBTTTG	TTGBTTT	9	9 TTGBTTT
14	Length table			12	6	B	R	R	T	G	B	B	R	BRRTGBBR	BRRTGB	6	6 BRRTGB
15	0	3	0.1	13	4	G	Y	T	Y	Y	Y	R	Y	GYTYYYRY	GYTY	3	3 GYTY
16	0.1	4	0.15	14	5	T	T	Y	Y	R	T	G	R	TTYYRTGR	TTYYR	5	5 TTYYR
17	0.25	5	0.2	15	7	Y	Y	T	B	Y	R	B	T	YYTBYRBT	YYTBYRB	8	8 YYTBYRB
18	0.45	6	0.3	16	3	T	R	B	Y	R	R	B	Y	TRBYRRBY	TRB	2	2 TRB
19	0.75	7	0.2	17	8	R	B	G	B	Y	R	Y	B	RBGBYRYB	RBGBYRYB	8	8 RBGBYRYB
20	0.95	8	0.05	18	6	Y	B	Y	T	T	Y	R	B	YBYTTYRB	YBYTTY	7	7 YBYTTY
21			1	19	5	Y	G	T	T	R	B	T	Y	YGTTRBTY	YGTTR	8	8 YGTTR
22				20	5	G	B	B	G	Y	T	G	B	GBBGYTGB	GBBGY	4	4 GBBGY
23																	

Length: E3: =VLOOKUP(RAND(),$A$15:$B$20,2)

Letters: F3: =VLOOKUP(RAND(),$A$3:$B$10,2)

Total: N3: =CONCATENATE(F3,G3,H3,I3,J3,K3,L3,M3)

String: O3: =LEFT(N3,E3)

Points: P3: =E3+RANDBETWEEN(-2,3)

Coded: Q3: =CONCATENATE(TEXT(P3,"#")," ",O3)

Sprint 1	Sprint 2	Sprint 3	Sprint 4
6 RYR	3 TGTR	3 RTT	7 GGTGGT
5 RGYGY	4 TRGRTT	5 GYYR	6 BRRRGR
3 BBBRB	9 YBRGYB	3 GGR	5 RGBTYG
4 YTRB	6 TBTYRY	8 RYTTBT	4 YRBGY
6 BBBRG	6 RBBGYBG	7 YBRGRR	8 TBRGRY
3 RBYGY	6 BTYBBYRG	8 BTRBBB	7 YTGRG
8 TRYTT	4 TBGBBB	2 YYGT	5 BBBB
10 TGGYTTG	8 TRGGY	6 BYYRGGT	10 RTGBYBTT
6 TYBGY	2 BRB	4 YYRYT	6 TYGT
5 GBRTTT	4 RBY	9 TGYYGT	8 RBTRBYG
3 RTG	8 YRRBRRTR	6 YTTTYTY	9 TRBGBR
4 TYGGYT	10 YYGYTTG	7 TGTGGGRR	6 BYTTG
5 BYB	4 TRRTGG	9 GTRRYR	6 RRGT
8 RBRYGY	4 YRRTT	5 YGTTGBR	5 RYBBTTT
2 BBR	3 TTY	6 TGRBBT	9 YGBYRYBR
4 BRYRBB	4 YGBGRG	2 YGYY	8 GGRGTT
5 RYYYB	5 GGRBYYR	9 TRTYGRY	6 TTRTTR
9 YRGYRG	7 TTBBYGB	4 BTYTG	3 YYGYT
8 YTTBTB	5 BTTYT	4 TRTB	4 BTB
4 YTY	3 RRY	10 YYBYBGY	6 BRGTB

REFERENCES

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., & Jeffries, R. (2001). *Manifesto for agile software development*.

Cardozo, E. S. F., Araújo Neto, J. B. F., Barza, A., França, A. C. C., & da Silva, F. Q. B. (2010). *SCRUM and Productivity in Software Projects: A Systematic Literature Review*. 1–4.

Dikert, K., Paasivaara, M., & Lassenius, C. (2016). Challenges and success factors for large-scale agile transformations: A systematic literature review. *Journal of Systems and Software*, *119*, 87–108.

Fernandes, J. M., & Sousa, S. M. (2010). PlayScrum - A card game to learn the scrum agile method. *2nd International Conference on Games and Virtual Worlds for Serious Applications, VS-GAMES 2010*, 52–59.

Florit, M. (2020). *No Title*. https://www.tastycupcakes.org/2020/04/ball-point-wfh-game/ Retrieved May 29, 2021.

Hof, S., Kropp, M., & Landolt, M. (2017). Use of Gamification to Teach Agile Values and Collaboration: A Multi-Week Scrum Simulation Project in an Undergraduate Software Engineering Course. *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, 323–328.

May, J., York, J., & Lending, D. (2016). Play ball: Bringing scrum into the classroom. *Journal of Information Systems Education*, *27*(2), 87–92.

Naik, N., Jenkins, P., & Newell, D. (2020). Learning Agile Scrum Methodology Using the Groupware Tool Trello Through Collaborative Working. In *Advances in Intelligent Systems and Computing* (Vol. 993). Springer International Publishing.

Paasivaara, M., Heikkilä, V., Lassenius, C., & Toivola, T. (2014). Teaching students scrum using LEGO blocks. *36th International Conference on Software Engineering, ICSE Companion 2014 - Proceedings*, 382–391.

Paz, I. (2018). *No Title*. https://www.linkedin.com/pulse/scrum-lego-simulation-agile-game-build-plane-ignacio-paz/ Retrieved May 29, 2021.

Rigby, D. K., Sutherland, J., & Takeuchi, H. (2016). The Secret History of Agile Innovation. *Harvard Business Review*, 4.

Sibona, C., Pourreza, S., & Hill, S. (2018). Origami: An active learning exercise for scrum project management. *Journal of Information Systems Education*, *29*(2), 105–116.

Steghöfer, J. P., Burden, H., Alahyari, H., & Haneberg, D. (2017). No silver brick: Opportunities and limitations of teaching Scrum with Lego workshops. *Journal of Systems and Software*, *131*, 230–247.

Sutherland, J. (2014). *Scrum: the art of doing twice the work in half the time*. Crown Business Publishing.

Sutherland, J., Patel, D., Casanave, C., Miller, J., & Hollowell, G. (2012). *Business object design and implementation: Oopsla'95 workshop proceedings 16 October 1995, Austin, Texas*. Springer Science & Business Media.

Takeuchi, H., & Nonaka, I. (1986). The new new product development game. *Harvard Business Review*, *64*(1), 137–146.

DECISION SCIENCES INSTITUTE

Requirement Analysis and System Design of a Smart Assistive System for Outpatient Care

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Pan Zou
Georgia Institute of Technology
Email: pzou3@gatech.edu

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

This work proposes the requirement analysis and system design of a smart assistive system (SAS) for outpatient care. The system aims to provide personalized health care service for patients out of the hospitals by using smart sensing, predictive learning, and intelligent reasoning. The proposed system can detect the abnormalities of the patient based on physiological data and provide assistive guidance based on the patient's case. With using the SAS, the cost and inconvenience can be reduced by providing personalized health services. The paper includes a case study of using SAS to provide service to an outpatient with heart arrhythmia.

KEYWORDS: Smart Sensing, Neural Network, Medicare accessories, Decision-making

INTRODUCTION

In 2020, collected from three thousand hospitals, around thirty million patients are discharged from the hospital (American Hospital Directory, 2020). It has also been reported that 19.6% of the outpatients were admitted by the hospital in 30 days in 2020 (Khera et al., 2019). Suboptimal communication of discharge instructions is one of the main factors that cause hospital readmission (Ashish et al., 2009). Despite the inappropriate discharge planning from the hospital side, a survey conducted in 2004 indicated that most outpatients have limited knowledge about the purpose, course, and side effects of their medication (Freyer et al., 2016). Failure in understanding the doctors' instructions could lead to the disease's reoccurrence and several complicated diseases. In addition, most outpatients have little access to their health conditions, the lack of approaches to personal monitoring health result in untreated or delayed-treated illness reoccurrence. Furthermore, the approach of accessing prompt treatment is limited to visiting the hospital for most outpatients. For instance, outpatients who have common health problems like diabetes, cardiovascular diseases, high blood glucose level, and high blood pressure are recommended to receive medical inspections periodically (Shahriyar et al., 2009).

Smart mobile devices bring the idea of the mobile health service system, which enables patients' high participation in their health. The mobile health system aims to provide personal continuous healthcare service to patients at an acceptable cost out of the hospital (Silva et al., 2015). There are a lot of services provided by the mobile health system in a healthy lifestyle, diet tracker, medication tracker, compliance, mental health, and children's health areas. For mobile health service systems which utilize the smart mobile device only, the provided services are limited. The development of Internet of Things (IoT) technology and smart sensing technology broaden the application area of mobile health. The health monitoring system utilizes the wireless sensor network, and smart mobile devices have become a huge research area. In such a system, sensors collect several types of biosignals from the patients, and the patients can access the monitoring results through the mobile device. Masses of researchers designed and tested the smart health monitoring system that addressed different types of chronic diseases. Further analysis through artificial intelligence is employed on the collected data to detect abnormalities or illnesses (Darshan et al., 2015). However, those proposed health monitoring systems cannot provide further assistive guidance related to detected abnormalities and illness, which still requires routine hospital visiting. This paper proposes the analysis and design of a patient-centered healthcare system for smart assistive system (SAS) for outpatient care to solve the problems. The system is designed to detect abnormal health status and assist the user by providing case-based health guidance. The objectives of the provided service are to reduce the necessary physical visit to the hospital and benefit the disease recoveries. The personal healthcare services provided by the SAS can reduce the cost and inconvenience to outpatients.

The proposed system design has four major technical challenges: health data acquisition, feature extraction, abnormal health status detection, and providing cased-based guidance. There are four challenges exists in the different step of the system running process.

(1) Health data acquisition is the fundamental step in the system running process. The challenges are mainly from two aspects which are sensor selection and sensor management. As a smart monitoring assistive application system, there is a huge limitation on the types of accessible physiological at home. Different types of physiological of the patients are necessary for detecting the health status. Selected sensors should collect types physiological as much as possible, which brings difficulties to the sensor selection. Managing such multiple sensor systems, balancing the trade-off between the sensor using duration and collected data precision is also challenging. Short sensor using duration can cause additional inconvenience to the user, and low data precision can affect the detection result of the whole system eventually.

(2) Health data feature extraction is the crucial step for successfully detecting the abnormal disease status. For each type of physiological data, different types of features can be extracted and be ready for analysis. To successfully detect the abnormal health status, extracted features of the physiological data should be able to reflect enough characteristics for determining one type of abnormal health status. Furthermore, the proposed system is designed to be an integrated smart assistive system for outpatient care which requires the system can identify and detect multiple types of abnormal health status. Each abnormal health status detection algorithm may need several types of features extracted from different types of physiological data. Therefore, selection in the extracted features and a huge amount of necessary feature extraction methods can cause huge difficulties to the design of the proposed system.

(3) Abnormal health status detection is one of the main services provided by the proposed system. This service is designed for providing generalized abnormal health status detection by using machine learning technology. The main challenges are the detection result accuracy and the

amount of time to train the machine learning model. A huge amount of dataset and training time is necessary for machine learning-based detection methods. However, the required amount of training data cannot be satisfied by the data collected from a single user. The required amount of time for training can also cause huge inconvenience to the user.

(4) Approach of providing case-based guidance is the most crucial assistive service the proposed system provided to the patients. Since the guidance is provided based on the detected health condition and the basic information of the patients, a method that can establish the connection between the existed case and related guidance is extremely important. In addition, the complexity of the health conditions of the patients makes it impossible for the database to store all the cased-based guidance. Therefore, the approach of finding new guidance based on the new case and its most similar existed case should be considered.

This paper proposes the smart assistive system (SAS) for outpatient care to monitor and predict the health condition of the user. The outpatients can access the detected results for possible disease symptoms or reoccurrence. Outpatients also receive assistive guidance for better recovery and simple diagnosis. The proposed system contains a smart sensing platform for data acquisition and sensor management. After collecting the physiological data from patients, a Recurrent neural network (RNN) model is utilized for the system to determine the abnormal health status. Once the kind of abnormal health status is detected, case-based reasoning (CBR) will provide guidance based on patients' status. The rest of the paper is organized as follows. Section two introduces the background of the related topics of the paper and work done by other researchers. Section three contains the use case and the functional process of the proposed system. Section four illustrates the design and systematical architecture of the sensing platform for data acquisition. Section five demonstrates the RNN based predictive learning model for abnormal health status detection. Section six introduces the approach of providing assistive guidance. The illustrative case study is included in section seven. Section eight discusses the proposed system and section nine makes the conclusion.

LITERATURE REVIEW

IoT Based Smart Health Monitoring Service System

Internet of Things (IoT) based health monitoring system is a continuous health monitoring system for the patients. With the wireless transmission and a portable sensor network system, the health monitor can bring convenience to the user. The IoT-based health monitoring system can increase the patients' accessibility to the health care service since this kind of system could be easily used at home. There is a lot of work done by different researchers in the IoT-based health monitor area. In 2016, researchers constructed an IoT system to detect abnormalities in the heart through ECG(Electrocardiography) signal (Penmasta & Reddy, 2016). The ECG monitor measures the real-time ECG signal and transfers it to the microcontroller for abnormities detection. Some system designs are focused on other abnormalities detection like obstructive sleep apnea (OSA). An IoT-based health monitoring system is proposed to detect the OSA through sensors and provide the location of the patient when the patient is in danger (Cai et al., 2019).

The health monitoring service system is not limited to physiological data acquisition. More services can be provided by such a system. An Android platform-based health service system is designed for chronic disease recovers by implementing wireless communication and medical sensor detection technology (Kong et al., 2016). The system monitors several parameter signs like ECG, blood pressure, and blood oxygen. The data is analyzed in the server and then is transferred to the Android-based platform. The health monitor service system can also provide visualized and speech feedback to the user. A smartphone-based health monitor software system, “Wellphone” is designed to correlate the physical activity of the user and physiological data measured from the user, which can detect the progression of chronic diseases (Moser & Smith, 2015). The system utilizes statistical analysis to the collected physiological data for abnormalities detection.

Intelligent Mobile Health Monitoring System (IMHMS) a smart health care system that is designed to monitor the patients’ health condition and provide medical feedback to patients (Shahriyar et al., 2009). The wearable biosensors collect the physiological data from the patients and transmit the data to the Patient Personal Home Server (PPHS) and Intelligent Medical Server (IMS). PPHS manages the data and sensors for patients’ access and data transmission. IMS is employed as a portal for the medical staff to examine the patient. The IMS provides feedback on patients’ health conditions based on the diagnosis of professional medical staff.

Machine Learning-Based Health Condition Detection

Machine learning algorithms have been commonly used in the health area for aiding in making diagnostic decisions. There are tons of different machine learning algorithms designed to detect different diseases of the patients.

As one of the most significant and complex organs, research in heart disease detection attracts the attention of masses of researchers. The physical structure and working condition of the heart could be shown by the ECG signals. The ECG has excellent potential in detecting abnormalities of the heart. By using the kNN classifier and wavelet transform, ten types of heart diseases are detectable (Saini et al., 2015). Another detection algorithm based on long short-term memory (LSTM) is designed for the continuous cardiac monitoring device (Saadatnejad et al., 2020). The LSTM-based ECG classification algorithm has the advantage in accuracy and speed, which is more suitable for the wearable ECG monitoring device. There are five detectable heartbeat classes in the LSTM-based algorithm.

Despite cardiovascular disease, many applications of machine learning algorithms are applied in other chronic disease diagnoses. The prediction model of the chronic disease diagnoses like diabetes, hepatic fibrosis, dementia prediction, lung cancer is developed quickly. The developed predictive model achieves high accuracy, which is around 60% to 80% (Battinei et al., 2020). Most predictive models require hospital laboratory data that are not available for home health monitoring which is not achievable for a home health assistance system.

Sensing Technology

Several crucial physiological data like heart rate, body temperature, respiration rate, blood pressure, pulse oxygenation, and blood glucose could reflect patients’ general health condition and some diseases’ symptoms. The new sensing technology improves the accuracy and portability of the sensors.

Electrocardiogram (ECG) is a commonly used technique that could monitor the heart rate non-invasively. It monitors the surface protein generated by the periods of electrical depolarization and repolarization during one cardiac cycle of the heart cell (Azad & Lemay, 2014). Wearable monitoring device employed this technique commonly exists in the market. The wrist is a suitable body position for wearing an ECG device that has the least affection to the patients. As for body temperature, different sensor placement positions affect the measurement. For the non-invasive temperature measure method, the relationship between skin temperature and body temperature is not stable since it is also affected by the evaporation of sweat and the contact between the sensor and the body. Estimation is the most used method for body temperature based on skin temperature. Researchers have come up with an accurate estimation method based on wrist skin temperature (Sim et al., 2016). Compared to heart rate and body temperature, there is a lack of mature technology for measuring blood pressure. An estimation-based wearable blood pressure measurement method was presented by using optical fiber technology. This non-invasive method measures the pulse wave signal generated in the cardiac cycle to estimate the blood pressure. In the pulse wave of blood flow, two types of waves are the percussion wave and the dicrotic wave. The time difference between two waves' peaks is defined as pulse transit time. Pulse transit time of the pulse waveform has an inverse relationship to blood pressure. A wristband blood pressure sensor system was invented in 2020. The system can launch a certain type of laser to the skin and perceive it by using an optical fiber unit to receive the information of the pulse. Then the blood pressure can be estimated from the pulse information. The whole system is wearable and can be made in the form of a wristband (Li et al., 2020).

There are several techniques that can monitor blood glucose non-invasively. However, the most non-invasive and non-sample-based technique cannot monitor the blood glucose level continuously. Researchers in South Korea present a wearable-band type non-invasive blood glucose monitoring device. The device utilizes Vis-NIR analysis to determine the blood glucose concentration. However, this method has some limitations, and the monitoring result of the kind sensor also needs to be validated in the future (Rachim & Chung, 2019). Other reliable non-invasive blood glucose monitoring devices are not continuous, which utilizes wave signals like light, ECG to detect the glucose concentration through the skin (Siddiqui et al., 2018). However, these devices cannot continuously monitor the glucose concentration, which requires patients' action several times per day.

There are various types of respiration rate sensors, which include acoustic-based methods, airflow-based methods, and motion-based sensors. Due to contact between the user and the device, the respiration rate detected by most mentioned methods is accurate from a clinical perspective (Al-Khalidi et al., 2014). One wearable type sensor utilizes Holo-Hilbert spectral analysis on the photoplethysmogram (PPG) signal from the wrist to detect the respiration rate. The device is in the shape of the wrist band, which returns accurate results in the experiments (Chang, 2018).

ANALYSIS AND DESIGN OF THE SMART MONITORING ASSISTIVE APPLICATION SYSTEM

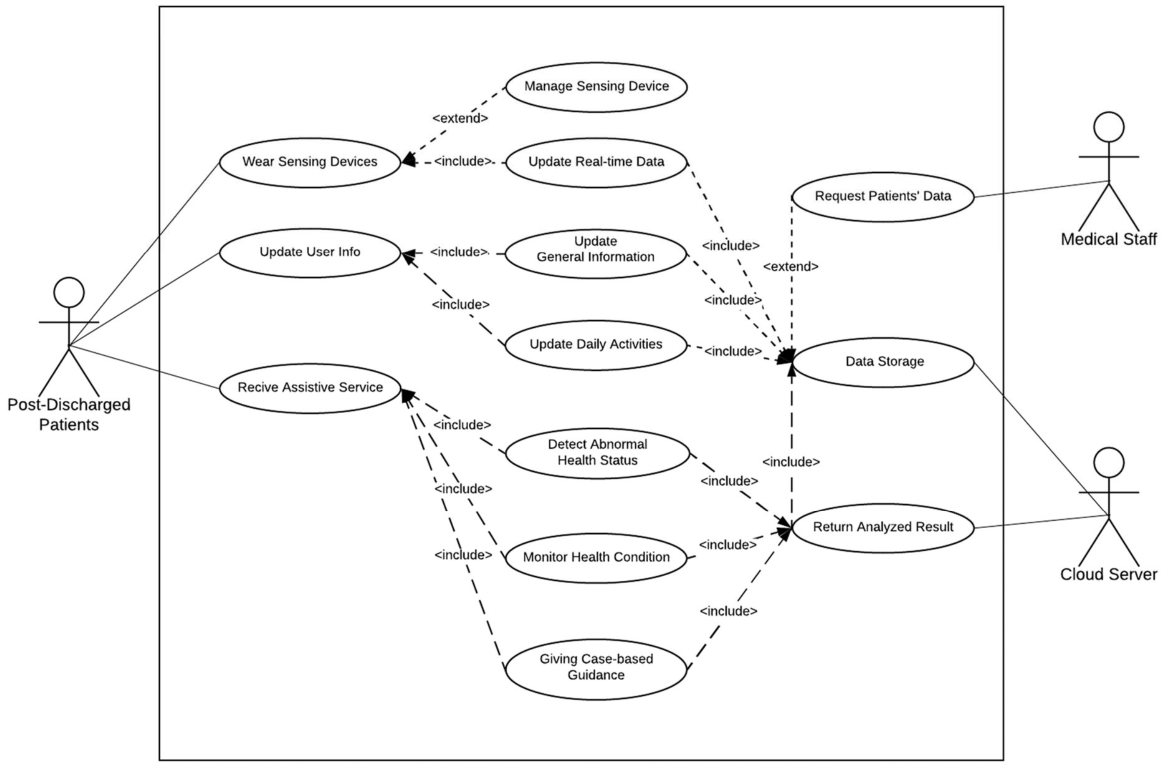
To solve the problems of outpatients. The proposed system is designed to have three main services. The first service is to detect the abnormal health status of the patients. The second is to provide feedback based on the health condition of the patients. The third is to monitor and store the physiological data of the user as information for doctors. The functional design of the whole system is concentrated on these three services.

Use Case Analysis

General use case analysis is discussed in this section firstly. The proposed system's services are mainly provided through the mobile platform-based application and the cloud server. The mobile platform act as a medium between the user, the doctor, and the decision-maker. The users are mainly outpatients, which is the primary actor in the system. They need follow-up health service, which includes abnormal status detection, notification, and cased-based guidance for the recovery. The doctor act as the secondary user who needs the long run collected of the user for diagnosis. The cloud server act as a decision-maker who analyses the collected data and returns corresponding prediction results. As the medium, the mobile manages the information flow of the whole system, which constructs the bridge between the user, the sensors, the doctor, and the cloud servers. A UML case diagram is shown in figure 1. Thus, in the UML case diagram of the SAS, the mobile platform is the container that encloses the interactions between three actors, the outpatients, the cloud server, and the medical staff. The outpatient is the primary user of the system.

To use the system, they first need to upload the user information to the mobile platform. There are two types of information they can upload to the platform. The first kind is their biographical information like sex, weight, and age to the mobile application to use the system. This type of information is categorized as general information in the UML case diagram. The second type of information is the daily activities of the user. The daily activities include the amount of exercise, food intake, and any other activities that could affect patients' health status. Compared to the general information, which only needs to be uploaded in a long time interval, the daily activities require the patient routinely. To reduce the amount of work to the user and let the decision-maker read the information directly, the daily activities are uploaded in terms of multiple choices and blank filling.

Figure 1: UML case diagram of the proposed system



After uploading the user information, the outpatients need to set up the equipment for data acquisition. The equipment of data acquisition contains sensors and a smart sensing platform. To measure different types of physiological data, there are different types of sensors. To collect the data continuously and reduce the inconvenience, most sensors are wearable for the patient. Non-wearable sensor requires the patients to take the measurement in a daily routine. As for the smart sensing platform, it is integrated with the mobile platform, which can be easily accessed and tuned by the user. The real-time collected data, general information, and daily activities are managed by the platform before uploading. The medical staff can access the managed data storage for physiological data records. The records can help the medical staff know about the health status of the patient.

The cloud server takes the uploaded data as input. By implementing the predictive learning model and the case-based reasoning model to the input, the cloud server will return the predictions and guidance to the patient. The predictions are the possibilities of the states of the abnormal health status. The guidance is based on the possible abnormal health status, users' daily activities, and users' general information. The health status controls the general trend of the guidance. Based on the daily activities, it can give more details to that trend and make it clear for the user. The doctor can also access the analyzed result to help them make the decision in diagnosis.

Functional Analysis

The function provided in the proposed system is based on the provided service. Figure 2 shows the IDEF0 diagram of the proposed system. The idef-0 diagram provides more detailed information about the procedure of the whole working flow of the proposed system. Different kinds of techniques are utilized in the different steps of the system, which include adaptive sensing, data processing, feature extraction, LSTM based predictive learning models, and case-based reasoning guidance providing a model. To begin with, the user will firstly choose sensors to use. The sensor selection is based on the patients' disease during the inpatient time, patients' intentions, and doctors' suggestions. This information belongs to the knowledge domain. Once the patients start using the system, it needs to use the sensors to collect the physiological data. At the same time, the patients are also required to update information. Both collected data and information are ready for the processing step. The processing of the data is at the cloud server. With certain processing procedures, the data can read by the feature extractor. Similar to the sensor selection, the feature extractor determines the types of extracted features based on patients' information. The extracted features act as the input of the pre-trained LSTM model. For each kind of abnormal health status, an LSTM-based model is created in the cloud server to make predictions. The pre-trained LSTM-based model could quickly finish the training and setting procedures based on the new information from the user. The result of abnormal health status is returned to the patient. Despite directly notifying the user, the result is also employed as the input for the guidance providing. Information updated by the patients is also required here for the CBR model to make the decision. The CBR model returns the guidance to the user to adjust its behavior for better recovery.

Figure 2. IDEF0 Diagram of the Overall Working Process of the System

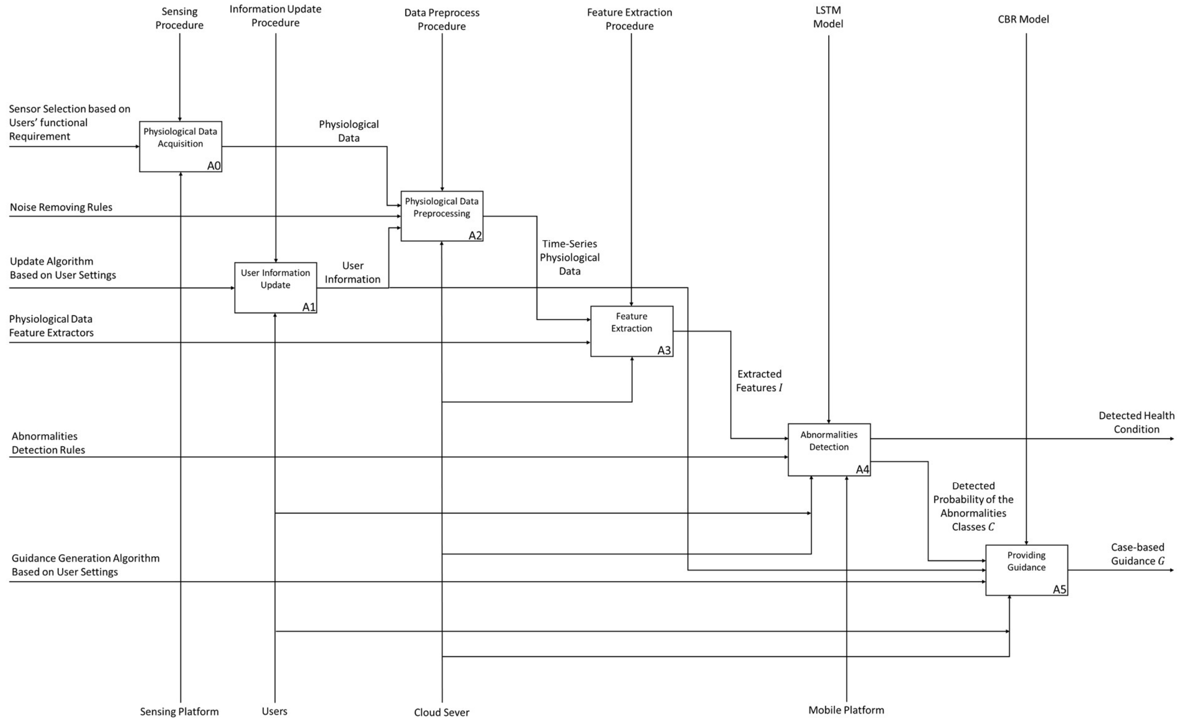
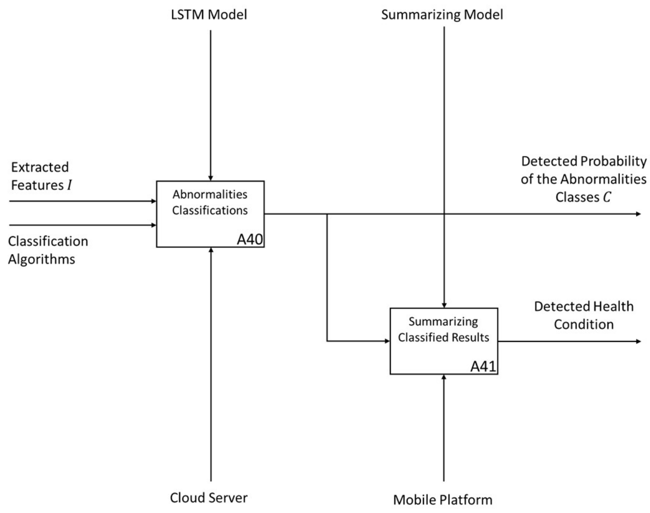


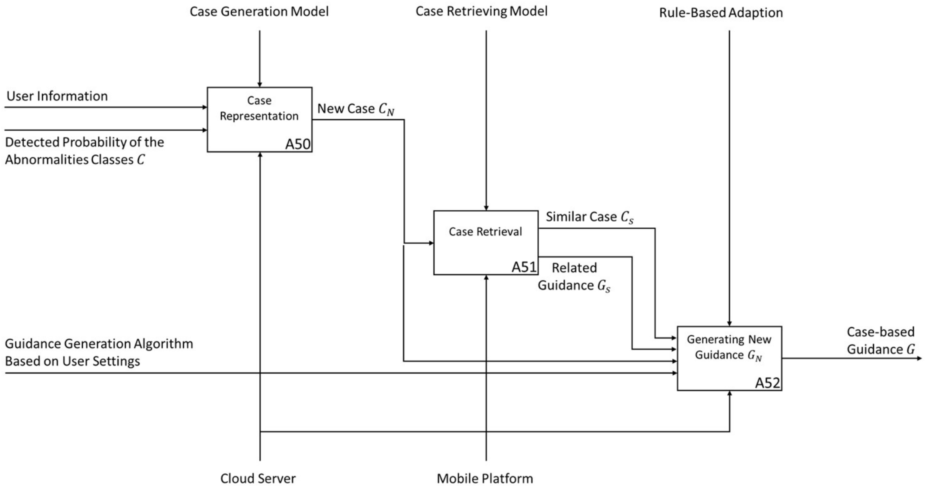
Figure 3 demonstrates the detailed workflow of the health abnormalities detection node A4. Extracted features $\boldsymbol{I}$ outputted from node A3, and the classification algorithms are input for the abnormalities classification node A40. The classification utilizes the LSTM model to find the possibility of each class of possible abnormalities $\boldsymbol{C}$ in the cloud server. The detected possibilities of each class of possible abnormalities $\boldsymbol{C}$ is one of the outputs of node A3. Node A41 summarizes the classified results based on the input $\boldsymbol{C}$ in the mobile platform. The summarized result is the detected health conditions which will be sent directly to the outpatient.

Figure 3. IDEF0 Diagram of the Node A4



The workflow inside providing guidance node A5 is shown in figure 4. There are three steps for providing the guidance: Case representation for generating the current case $\boldsymbol{C_N}$; Case retrieval for finding the most similar case and related guidance to the current case; Generating new guidance $\boldsymbol{G_N}$ for the current case. The new case $\boldsymbol{C_N}$ outputted from node A50 goes into the case retrieval. By implementing the case retrieving model in the cloud server, The most similar case $\boldsymbol{C_s}$ and related guidance $\boldsymbol{G_s}$ are found and sent to node A52 to generate the new guidance based on the rule-based adaption.

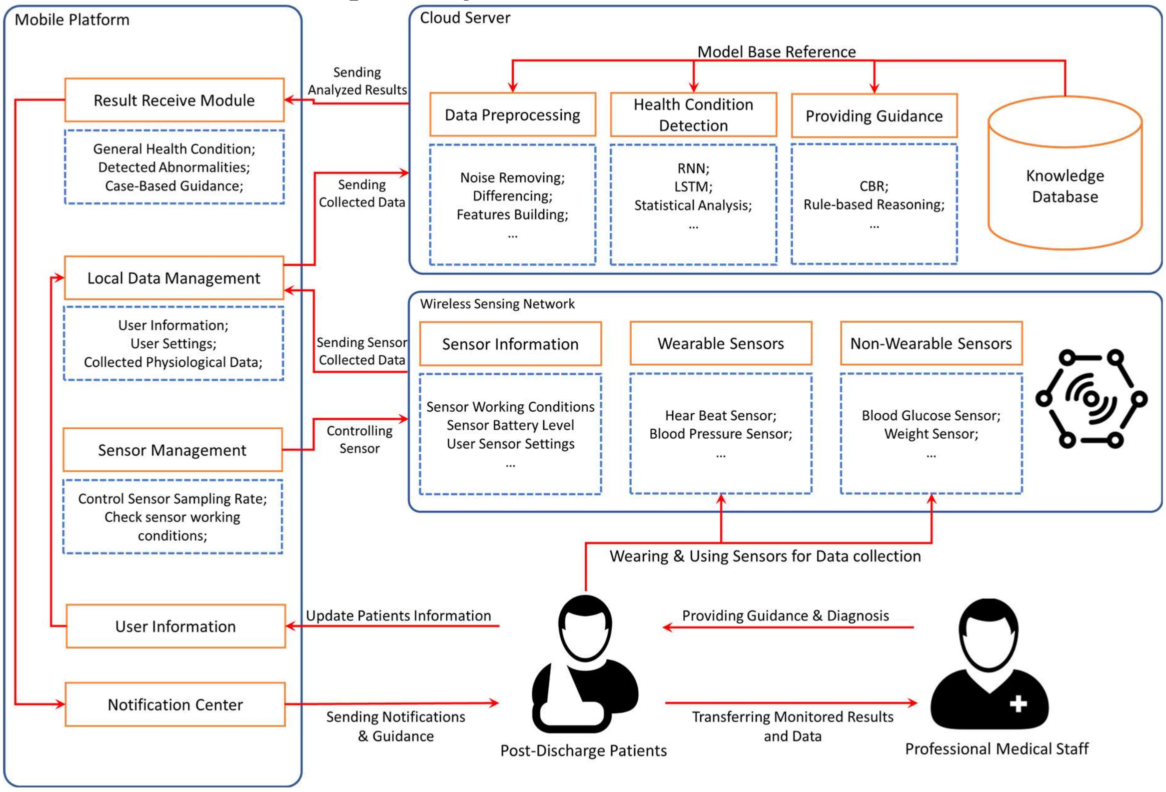
Figure 4. IDEF0 Diagram of the Node A5



System Architecture of the smart monitoring assistive application system

Figure 5 shows the system architecture of the proposed smart monitoring assistive application system. There are five components: Mobile Platform for notification, sensor management, and data transfer; Cloud server for data preprocessing, running abnormalities detection algorithms and running reasoning model for providing case-based guidance; Wireless sensing network for collecting physiological data from the user; Outpatients who are the main users of the proposed system; Professional medical Staff for diagnosis based on the monitored results of the system.

Figure 5: System Architecture of the SAS



To demonstrate the system architecture, an illustrative example is given. A male outpatient who has cardiovascular disease uses the proposed system. Besides cardiovascular disease, the patient is overweighted, which indicates he may have potential health problems with high blood pressure. The patient aims to use the system to monitor his heartbeats, blood pressure and receive notifications when an abnormality is detected by the system. The proposed system also provides guidance to benefit his cardiovascular disease recovery. To collect the physiological data from the patient, two types of wearable sensors, the heartbeats sensors, blood pressure sensors, are used by the patient. Those sensors collect the ECG signal from the patient, which contains the heartbeat and blood pressure information. The sensors construct the wireless sensing network, which collects and transfer the physiological data (ECG signal of the heartbeat and blood pressure) to the mobile platform. The sensor management module in the mobile platform is designed to control the sensor in terms of the sampling rate. For example, when the battery level of the heartbeat sensor is low, the time interval of collecting heartbeat ECG signal data will be larger to reduce energy consumption. This will be discussed in section 6 in detail. To gather the information of the sensors, the sensor information of each sensor is also transferred to the mobile platform for sensor management, and the information includes the battery level of the sensors, sensor working conditions, and user sensor settings. The patient also updates his personal information to the mobile platform includes his sex, age, height, weight, and the use purpose of the system. Collected heartbeat ECG signal, blood pressure ECG signal, and personal information are temporarily stored in the mobile platform. Local data management transfers the temporarily stored data of the outpatient to the cloud server.

The server and database preprocess the time-series ECG signal of the heartbeat and blood pressure by reducing the noise and extract the feature of the data. Algorithms and procedures of the heartbeat and blood pressure ECG signal preprocess are stored in the knowledge database. Once the cloud server receives the personal information of the patient, preprocessing algorithms and procedures of the heartbeat ECG data and blood pressure ECG data are selected. The knowledge database also stores the abnormalities detection models for different types of the physiological signal. In this example, the detection model of heart abnormalities based on the heartbeat ECG signal and the detection model of the blood pressure based on the blood pressure ECG signal is processed in the server. The server also runs the model for providing guidance which is discussed in section 6. Results obtained by the server are sent to the mobile platform, which includes the possibilities of the abnormalities classification of the heart diseases, blood pressure, and case-based guidance. The notification center in the mobile platform returns the result to the patient. The patient could send the monitored data and results to the medical staff for further guidance or diagnosis.

SMART SENSING PLATFORM FOR HEALTH DATA ACQUISITION

Health data acquisition acts as the fundamental part of the proposed whole system. All three basic functions of the proposed system condition monitoring, symptoms detection, and assistive action require the physiological data from the patient. To detect the abnormal health status of the patient accurately, there should be a certain number of sensors for acquiring different types of data. Furthermore, real-time data from sensors causes huge storage and accessibility problems in the long run. Therefore, A smart sensing platform is necessary for managing wearable sensors, collecting physiological data, and processing the data. In the past several decades, development in sensing technology has provided masses of portable and flexible methods of collecting physiological data from patients, which make an integral smart sensing platform possible. The smart sensing platform has two crucial parts overall, which are sensors and the platform. The requirement of sensors is strict. Firstly, most sensors should be wearable and non-invasive to avoid causing inconvenience to patients. Secondly, the sensor should not be sample-based either. Thirdly, the combination of types of the sensor should reflect the symptoms of certain diseases. Balancing between the number and sensors and the types of detectable symptoms is extraordinarily challenging. The sizes of the sensors also affect the convenience of using the system. The platform is the second crucial part, which acts as a signal receiver and an assistive action achiever between the sensors and the sever. The platform should have a physical form to provide feedback to the patient. However, as same as the sensors, the physical shape of the platform should necessarily small.

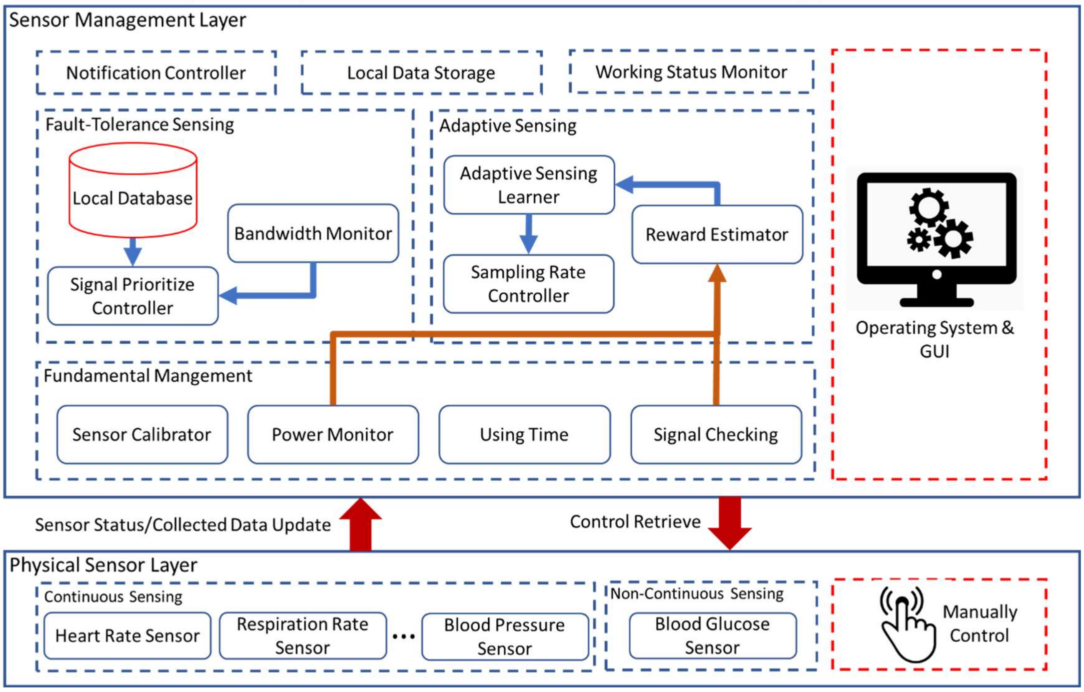
Design of the Wireless Sensing Network

The wearable sensors, non-wearable sensors, and the smart platform could be connected to form a wireless sensing network (WSN). In the WSN, both sensors and the smart sensing platform interact with the user. The sensor management is taken charge of by the mobile platform, which will be discussed in detail in this section. The mobile platform also returns the readings of the sensor and some sensor status to the user, which also allows the user to change some settings of the whole smart sensing platform.

Development in IoT and information networks has made the communication between platforms and sensors a mature technology. Several wireless communication technologies like Wi-Fi and Bluetooth could undertake a huge load of data transmission. Despite the hardware issue, there are two main technical problems, which are the energy usage of the sensor and the fault tolerance of the data communication (Panangadan et al., 2005). In order to continuously monitor the physiological data of the patients, the sensors are required to keep running for a long time. Since most of the sensors must be portable to worn by patients, the energy cost of the sensor is a vital issue. Fault tolerance is another vital challenge since the corruption of the physiological data transmission due to bad bandwidth could severely affect other services provided by the system.

System Architecture of Smart Sensing Platform

Figure 6: System Architecture of the Smart Sensing platform



The smart sensing platform is designed to overcome the challenges. As shown in figure 6, there are two main layers in the platform: the physical sensor layer and the sensor management layer. The physical sensor layer contains continuous sensors and non-sensors. The patient wears the sensor for continuous data monitoring and uses the non-continuous sensor to measure some physiological data. Two different types of sensors use two types of control and sensing algorithms. Most functions in the sensor management layers are designed for the continuous sensor like fault-tolerance sensing and adaptive sensing due to the data stream. Sensors in the physical sensor layers communicate with sensor management layers, which update the data and working status and receive the control order. In the fundamental management service, the sensor calibrator lets the user and platform calibrate the sensor. Using the time function reflects the run time of the system and sensors. The power monitor function monitors the power usage of the sensors by collecting the remaining power status from the sensor. The signal quality checking function checks the quality of collected data at the different sampling rates. These two functions are relatively important for adaptive sensing service in the sensor management layer, which will be detailly discussed while introducing the adaptive sensing service. The local data storage can store the setting and the physiological data for data resampling. It can be treated as a relay station for the temporary physiological data. Notification controller manages several kinds of notifications when the working status monitor detects a malfunction of the hardware working status.

Fault-tolerance service is designed to minimize the effect of corrupted data in low bandwidth conditions. It firstly ranks the different types of sensor collected data based on the priority. When the bandwidth becomes unstable or low, the signal prioritize controller lets the sensor management layer receive the more important type of physiological data first. Therefore, the more important physiological data has less chance to contain an error. The priority of the data type is stored in the local database. Based on different sensor types and different settings, the priority of different types of data can be altered. Settings like sensor data sampling rate can affect the size of the data stream, which makes the situation more complex.

Adaptive sensing service is one of the most important services in the sensor management layer. The sampling rate controlled by this service determines the size of the data stream, quality of the signal and sensor using time.

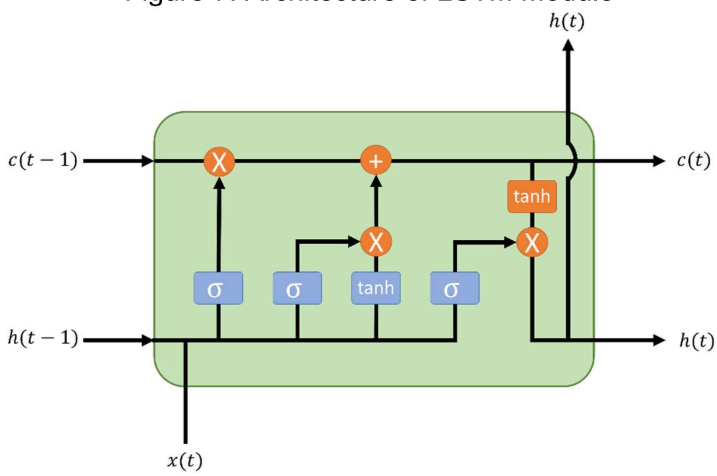
To manage multiple data flows, functions, and services in the sensor management layer, an embedded operating system is necessary. Some settings and the working status of the operating system can be shown and altered in the GUI by the user for convenience. However, the interaction between the users and sensors is limited to turn on or off manually.

DATA ANALYSIS AND PREDICTION LEARNING FOR CONDITION MONITORING AND SYMPTOM DETECTION

After collecting and temporarily storing the physiological data in the smart sensing platform, the data is transferred to the cloud server for processing. The algorithms in the cloud server process and analyze the data to extract features contained in the data. To detect the abnormal health status based on limited data, the predictive learning model RNN is employed for extracting the features. In this section, an LSTM-based abnormal heath status detection model is proposed.

LSTM Network Architecture

Figure 7: Architecture of LSTM Module



Compared to most neural networks, RNN contains feed connection in the neural net architecture (Medsker&Jain, 2001). The closing loop in the neural architecture lets the information flow persist, which allows the neural network to handle long terms of information. Compared to the traditional Convolution neural network (CNN), RNN is more capable of learning long-term dependencies data. Long Short-Term Memory networks (LSTMs) is one type of RNN model which is good at solving a sequence classification problem.

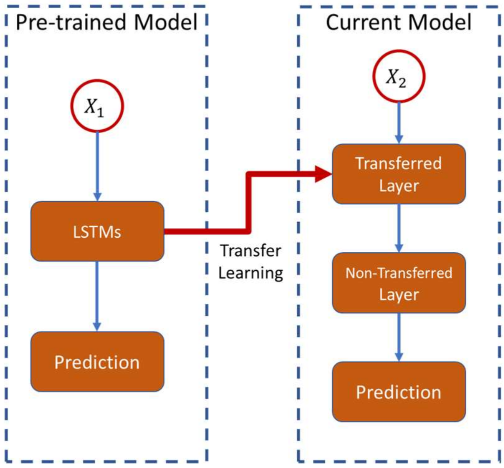
In a simple RNN model, modules of the neural network are repeated, the input training data at different steps goes into the layer in each repeated module to get a corresponded output. LSTM network is one type of RNN model which is much faster and more accurate than the standard RNN (Graves et al., 2005). The architecture of the repeating LSTM module is shown in figure 7. There are four neural network layers in the module. The input vector runs through the repeated cells, and four neural network layers interact with each other, which decide the dependencies of the information stored in the old or new cell states (Sherstinsky, 2020).

To train an LSTMs model, the physiological data collected by the sensors should be extracted for features. Since there are several different types of data collected by sensors, different kinds of features are required to be extracted from each type of physiological data. To detect the abnormal health status, a combination of the extracted feature from different physiological data is needed for each kind of symptom or health status. Furthermore, a different LSTM-based model is needed for a different type of abnormal health status.

Pre-trained RNN Model by Using Transfer learning

The training of the RNN model requires a certain amount of data, and the training process is also time-consuming. Therefore, to reduce the amount of setting time for the user, using a strategy like transfer learning with a pre-trained RNN model is the optimal choice.

Figure 8: Transfer Learning with Pre-trained Model



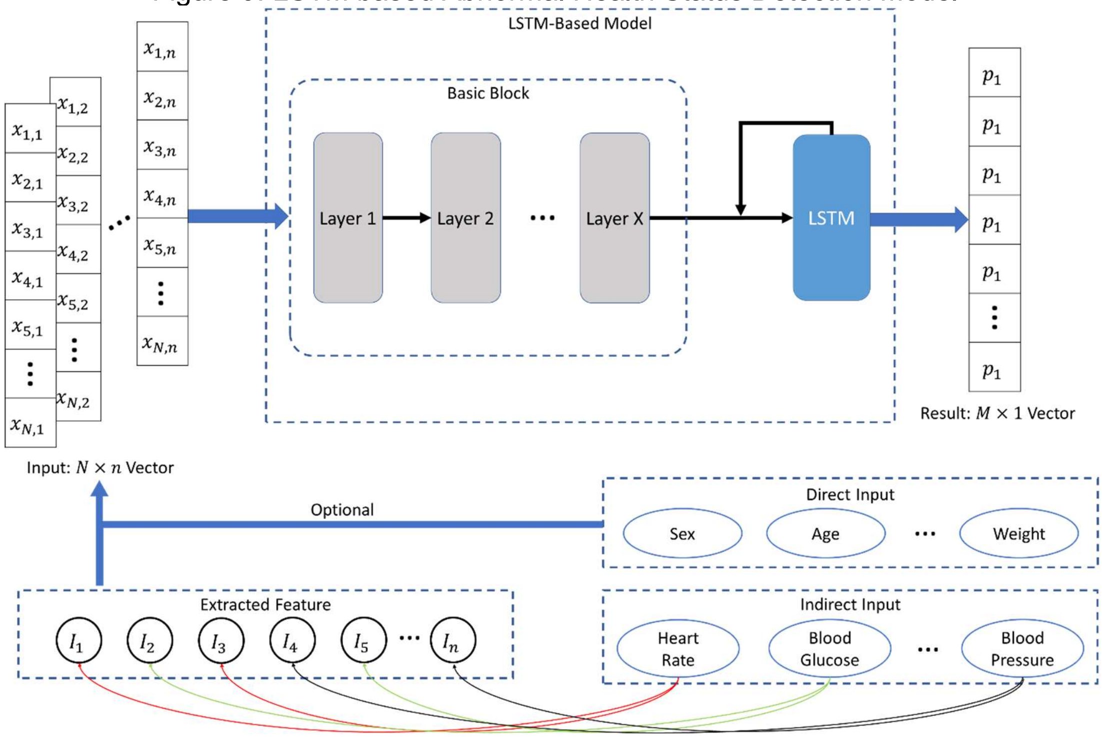
Since the predictive learning module is designed to be executed in the cloud server, parameter settings in the pre-trained model can be easily employed for the patients' individual model. As shown in figure 8, during the training process of the new model for patients, some of the layers in the pre-trained model are transferred to the new model. Other parameters of the layer are set to be default in the new model. Shorter training time can be obtained by using this approach.

Recurrent Neural Network Model for Abnormal Status Detection

To employ the LSTM model to detect abnormal health status, despite the physiological data collected from the sensor, some basic information of the user is necessary. The basic information like weight, age, and other general information requires manual input from the user. Different from sensor collected data, this information can be directly inputted into the LSTM model. Therefore, in the proposed LSTM model, this kind of information is also called direct input, as shown in figure 9. The direct input combines with the features extracted from the indirect input forms a vector as the input. The vector-formed input is $\boldsymbol{N} \times \boldsymbol{n}$. $\boldsymbol{N}$ is the length of the input vector, and $\boldsymbol{n}$ is the number of features that are used for the model. The input goes through the layers, and the trained LSTM model predicts the vector-formed result for one type of abnormal health status. The dimension of the result is $\boldsymbol{M} \times \mathbf{1}$ where $\boldsymbol{M}$ denotes the number of the possible state of that type of abnormal health status.

A simplified example could demonstrate the use of the model, and if the user chooses to use the system for detecting arrhythmia, the pre-trained LSTM model will be loaded in the cloud server and connected to the smart sensing platform. Heart rate data will be collected from the user. After executing the feature extraction process on the heart rate data, a $\boldsymbol{N} \times \boldsymbol{n}$ vector-formed input is formed, which contained features and some other manually input information. The LSTM model will process the input and predict the result, which is a $\boldsymbol{M} \times \mathbf{1}$ vector under the condition that we assume there are $\boldsymbol{M}$ kinds of states in arrhythmia.

Figure 9: LSTM based Abnormal Health Status Detection Model

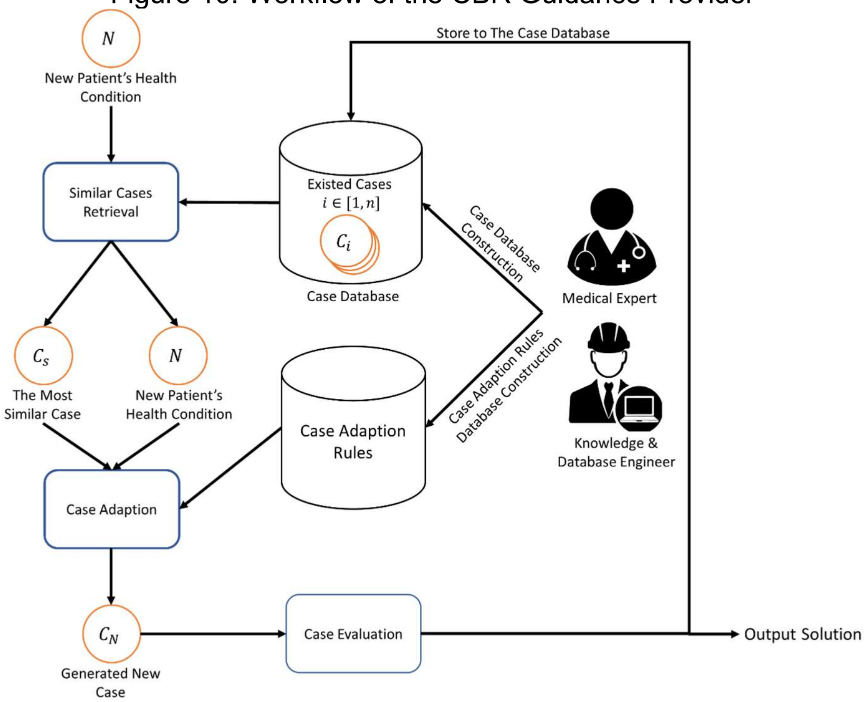


CASE-BASED REASONING GUIDANCE PROVIDER

To provide case-based guidance to the user, the proposed system utilizes the CBR approach to evaluate cases and optimize the guidance. CBR is a technology of generating new solutions to the new case based on the existed old solution (Kolodner, 1988). The Cased-based Reasoning Guidance Provider (CBRGP) is provided and designed to generate new assistive guidance to the new patient's health condition and based on finding the most similar existed patient's case. The new health condition of the patient and its generated solution can form a new case to be stored in the database.

The workflow of the Case-based Reasoning Guidance Provider is shown in figure 10. The use of CBRGP heavily relies on the existed case database and adaption rule database. By using technology like data mining, rules evaluation, and classification, medical experts and database engineering can establish the existed cases database and case adaption rules database with a raw assistive guidance database (Huang et al., 2007). After acquiring the new health condition $\boldsymbol{N}$ from the user, the CBRGP retrieves the most similar case stored in the case database. The case database stores the existed case $\boldsymbol{C_i}, \boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, .., \boldsymbol{n}$. After retrieving the most similar case $\boldsymbol{C_s}$, the CBRGP adapt or reuse the assistive guidance or treatment stored in the most similar case $\boldsymbol{C_s}$ to generate the new assisted case $\boldsymbol{C_N}$. The new assisted case $\boldsymbol{C_N}$ is evaluated in the CBRGP, once it passes the evaluation, the assistive guidance is outputted to the user, and the new assisted case $\boldsymbol{C_N}$ is stored in the case database.

Figure 10: Workflow of the CBR Guidance Provider



The existed patient's case contains: (1) health information which contains the physiological data, basic information, and abnormalities condition detected by the proposed LSTM-based system, (2) assistive guidance in the forms of natural languages. The natural language-based assistive guidance has a substitution part that is changed through the adaption process.

Similar Patient's Case Retrieval

To evaluate an existing guidance solution and adapt it to the new case, the case that is the most similar to the new case and its solution should be retrieved by the system. After the general filtering process, the information of a set of selected similar cases is returned to the CBR guidance provider for further evaluation.

The patient case is represented as a $\boldsymbol{N}$-tuple: $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_{m-1}}, \boldsymbol{v_m}, \boldsymbol{p_{m+1}}, \ldots, \boldsymbol{p_n}\}$. For the $\boldsymbol{1_{st}}$ to $\boldsymbol{m_{th}}$ tuple, they store the value $\boldsymbol{v}$ inputted by the user like height and weight. These are real number values. For the $(\boldsymbol{m}+\boldsymbol{1})_{\boldsymbol{th}}$ to $\boldsymbol{n}$ tuple, the possibility $\boldsymbol{p}$ of the detectable abnormalities. Due to the huge differences between the value of inputted information and the value of possibilities, the tuple value is rearranged into the interval of $[\boldsymbol{0}, \boldsymbol{1}]$ for calculation. The comparison function between the two cases is shown below:

$$\boldsymbol{f}(\boldsymbol{C_i}, \boldsymbol{C_j}) = \sum_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{m}} \boldsymbol{w_k}(\boldsymbol{v_k^i} - \boldsymbol{v_k^j}) + \sum_{\boldsymbol{l}=\boldsymbol{m}+\boldsymbol{1}}^{\boldsymbol{n}} \boldsymbol{w_l}(\boldsymbol{p_l^i} - \boldsymbol{p_l^j}) \tag{5}$$

The $\boldsymbol{C_i}$ and $\boldsymbol{C_j}$ are the two compared cases, the indexing of $\boldsymbol{i}$ and $\boldsymbol{j}$ in the equation show which tuple that the value belongs to. In order to find the most similar case,

$$\mathbf{max}(\boldsymbol{f}(\boldsymbol{C_n}, \boldsymbol{C_s})), \boldsymbol{s} \in \boldsymbol{S} \tag{6}$$

A similar case could be retrieved by using the objective equation, which finding the biggest similarity score between the selected cases $\boldsymbol{C_s}$ and the new case $\boldsymbol{C_n}$. The index $\boldsymbol{s}$ belongs to the subset of selected case $\boldsymbol{S}$. After retrieving the most similar cases, the corresponding guidance is also retrieved by maximizing the objective function.

New Guidance Generation

The guidance corresponding to the most similar case and the difference between the new case and its most similar case is used for the new guidance generation. The general new guidance generation process contains five basic steps (Kolodner, 2010):

(1) Substitution is the first step of the new guidance generation process. The invalid part in the old guidance is replaced by the new material based on the difference between the new case and its most similar case. It also adjusts the parameter in the old guidance.

(2) Adaption is the most crucial step of the new guidance generation process. After identifying the parameter type in the old guidance that needs to be fixed, a rule-based adaption strategy is utilized to refine the new guidance to the new case.

(3) Evaluation is to judge the generated guidance based on the future health condition of the patient and the professional medical staff.

(4) Update is the last step which is executed after the successful generation of the new guidance and positive feedback of the evaluation. The new guidance is updated to the database with its new corresponding case for future use.

A CASE STUDY OF THE SAS FOR OUTPATIENT CARE WITH HEART ARRHYTHMIA

This section presents the case study of using the SAS. The case demonstrates the use of the proposed system. In the case study, the proposed system provides service to a patient who had heart arrhythmia problems. The rest of the section includes hardware setup, heart arrhythmia detection, case representation, and a similar case retrieve process in the proposed system.

Case Description

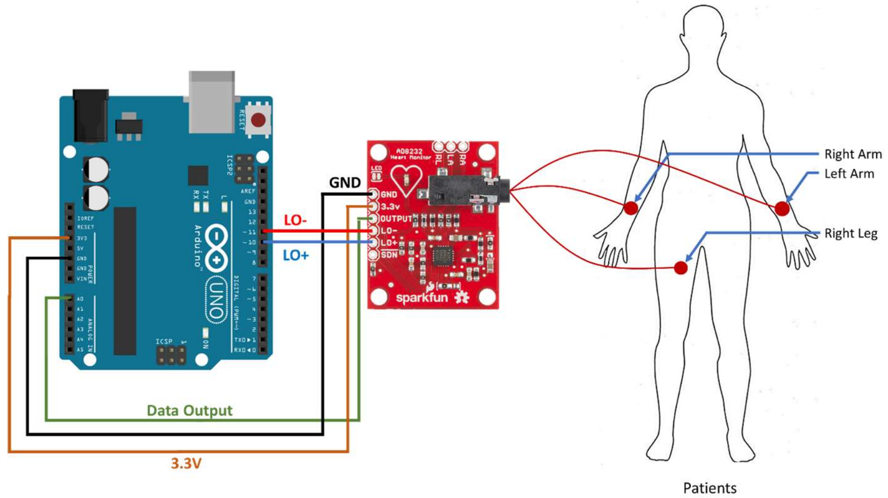
Heart Arrhythmia is one of the major health problems in the world. Heart arrhythmia refers to irregular electrical heartbeat signals, which include single ectopic occurrence and continuous arrhythmia (Hosseini et al., 2001). ECG heartbeat signal is crucial for cardiovascular disorders (CVDs) diagnosis, which can be collected by ECG sensor simply and non-invasively. In the case study, the user is a 60 years-old male who is diagnosed with heart failure. The outpatient decides to use the proposed system for the assistance of the monitor and recovery of heart diseases.

The outpatient wears an ECG sensor for heartbeat data collection. The collected time-series heartbeats signal is preprocessed and then classified by the pre-trained LSTM-based model. A new case of the patient is generated by the classification result of a time interval and uses information. A similar case stored in the database is retrieved for generating the guidance to the new case.

Hardware Setup of the Heartbeat Monitor

The hardware setup of the case study is shown in figure 11. The system consists of two parts, an Arduino Uno board employed as the microcontroller and an AD8232 ECG Analog sensor. The electrode pads are attached to the right arm, left arm, and right leg and connected to the input port of the sensor board.

Figure 11: Hardware Setup of the ECG Heartbeats Sensor

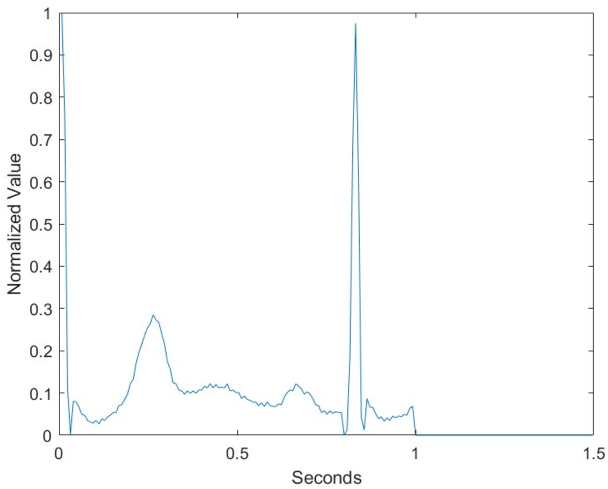


The Leads-off Detect – pin and Leads-off Detect + pin of the AD8232 are connected to pin 10 and pin 11 of the Arduino Uno board. These pins are employed for lead-off detection. The ECG signal detected by the sensor is transmitted through the output port A0 pin on the Arduino.

LSTM-based Heart Arrhythmia Classification Model

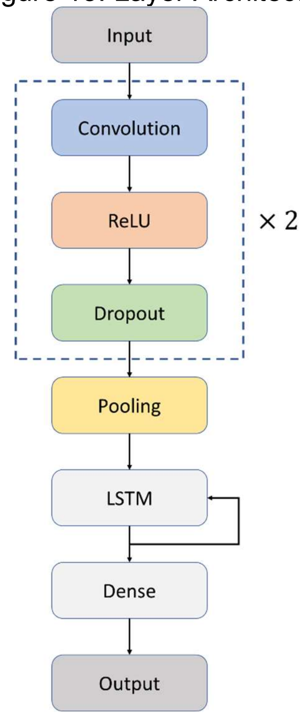
The ECG signal is collected by the sensor for heart anomalies detection. The model employed for classifying the ECG signal is discussed in this section. The data for training the model is originally combined from MIT-BIH Arrhythmia Database (Amaral et al., 2000) and PTB Diagnostic ECG Database (Kreiseler et al., 1995). The dataset is resampled at 125Hz and labeled with five different beat categories which are ready for training the LSTM-based heart arrhythmia classification model directly. (Kachuee et al., 2018). The ECG heartbeat data have five categories: $\boldsymbol{N}, \boldsymbol{S}, \boldsymbol{V}, \boldsymbol{F}, \boldsymbol{Q}$. Among the five categories, $\boldsymbol{N}$ represents the normal beats, $\boldsymbol{S}$ represents the supraventricular ectopic beats, $\boldsymbol{F}$ represents the fusion beats and $\boldsymbol{Q}$ represents the unknown beats. The long period of the ECG heartbeat signal is normalized and divided into a short period of 1second based on the R-R time interval. An example of a divided ECG heartbeat is shown in figure 12.

Figure 12: Divided ECG Heartbeat Signal



To train the model, the data split into 75% of the training data and 25% of the testing data. There are 87554 samples of training data and 21892 samples of test data. During the training process, 20% of the training data was employed as the validation data. Each short period of the ECG heartbeat signal and its derivative are the input of the LSTM-based classification model. The size of the input is $\mathbf{187} \times \mathbf{2}$, where the first row contains the original ECG heartbeat signal, and the second row contains the derivative. The layered architecture of the model is shown in figure 13. The input data goes through the block that contains a 2-D convolution layer, a ReLU activation layer, and a dropout layer twice. The dimension of the feature map is reduced through an average pooling layer. Then, the output data goes into the LSTM layer, and a softmax layer returns the output of the model. The results of classification possibilities of five classes.

Figure 13: Layer Architecture of the LSTM-based Neural Network



Input	187 × 2
Conv_1	36 × 1,256
Re_lu_1	36 × 1,256
Conv_2	32 × 1,512
Re_lu_2	32 × 1,512
Pooling	16 × 1,512 Average
LSTM	64
Dense	5 Classes

Figure 14 shows the training curve of the model. The training contains 200 epochs. The overall training accuracy reaches the max value of 99.77%, and the overall validation accuracy reaches the max value of 98.89%. The results show the effectiveness and reliability of the model.

Figure 14: Training Curve of the Model Training Process

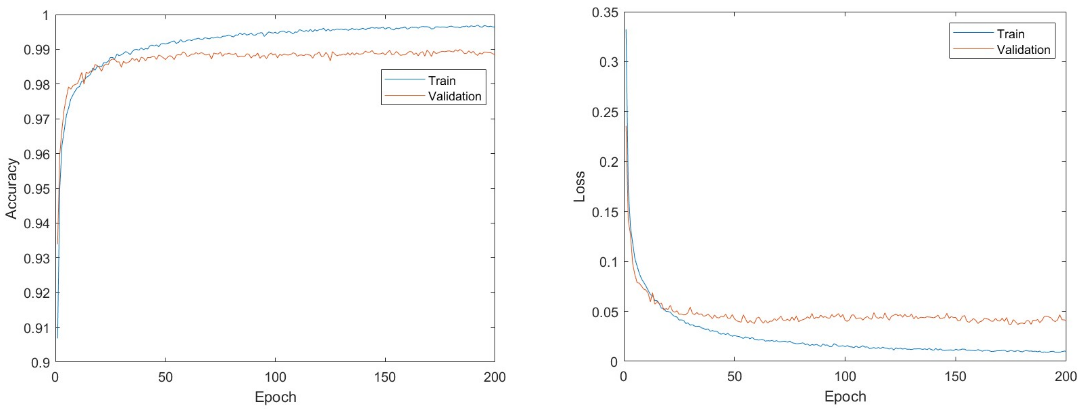


Table 1 shows the experimental results of heart arrhythmia classification. The prediction results are 99% accuracy for the normal beats, 84% accuracy of the supraventricular ectopic beats, 96% accuracy for the ventricular ectopic beats, and 99 % for the unknown beats. The prediction accuracy of detecting supraventricular ectopic beats and fusion beats is relatively low. Considering that the input signal is around one cycle of the ECG heartbeat signal, the possibility of detecting the anomalies class wrong is little since a large amount of ECG heartbeat data can be provided.

Table 1. Experimental Results of Heart Arrhythmia Classification

	Normal Beats	Suparventrical Ectopic Beats	Ventricular Ectopic Beats	Fusion Beats	Unknown Beats
Accuracy	0.99	0.84	0.96	0.86	0.99

Case-based Reasoning Assistive Guidance Provider

The cased-based Reasoning model is employed for providing assistive guidance to the user after abnormalities detection. Since the LSTM-based classification model takes the single heartbeat as the input, therefore the classification result of one single heartbeat can't reflect the general heart condition of the user. Therefore, the detection results of the heartbeats in an hour were averaged before applying the CBR mechanism.

Table 2. Overall Classification Results of Heartbeats of the User in Two Hours

Predicting Results of the Arrhythmia Type Classification
Probability of Normal Beats (N): 0.20851
Probability of Supraventricular Ectopic Beats (S): 0.60245
Probability of Ventricular Ectopic Beats (V): 0.14902
Probability of Fusion Beats (F): 0.39987
Probability of Unknown Beats (Q): 0

Table 3. Results of the CBR Assistive Guidance Provider

Name: John Smith Sex: Male Age:65	Weight:75.1 kg Height: 1.63m
Similar Case Number:	4
Similarity checking items:	(1) Age (2) BMI (3) Probability of Normal Beats (N) (4) Probability of Supraventricular Ectopic Beats (S) (5) Probability of Ventricular Ectopic Beats (V) (6) Probability of Fusion Beats (F)
Similarity:	72.82%
Assistive Guidance:	(1) Controlling the weight (2) Managing diet and reduce caffeine and alcohols intake (3) Do low aerobic exercise

The detection result of heartbeats in two hours of the user is shown in table 2. Supraventricular Ectopic Beats has the highest probability of 60.245%. The probability of normal beats, ventricular ectopic beats, and fusion beats are 20.851%, 14.902%, and 39.987%. Combining with user input information of sex, age, weight, and height, the new case of the patient is generated and sent to the CBR assistive guidance provider. Table 4 shows the results of the CBR assistive guidance provider.

In this case study, the CBR assistive guidance provider has four similarities checking number, which are (1) Age, (2) BMI (calculated from weight and height), (3) Probability of Normal Beats (N), (4) Probability of Ventricular Ectopic Beats (V), and (6) Probability of Fusion Beats. By checking the similarity, the most similar case is No. 4 with a similarity of 72.82%. The new guidance to the new case is generated by the guidance adaption. The assistive guidance to the new case is also given in table 3. The SAS suggests the patient (1) control the weight, (2) Managing diet and reduce caffeine and alcohol intake, (3) Do low aerobic excises.

DISCUSSION

The case study presents the general working flow of the proposed system for a outpatient who has a heart arrhythmia. The outpatient wears ECG heartbeat signal sensors and uploads his basic information to the system. The physiological data collected by the ECG heartbeat sensors is sent to the system for abnormalities detection. The ECG heart signal is preprocessed where QT intervals in the heartbeat signal are identified and segmented. The first derivative of the segmented time-series ECG heartbeat signal is also inputted to the classification model with the segmented ECG signal. An LSTM-based neural network model is employed for classifying the type of abnormalities with high accuracy and short time. The classified result and other users' information is inputted into the CBRGP system, then the assistive guidance is generated based on the adaption to the retrieved most similar case.

The proposed SAS has the potential to reduce the cost and difficulties of post-discharge care. It also can assist the diagnosis process for the patient. However, the whole system relies on the construction of masses of databases which include the database stored different types of the abnormalities classification model, the database that stores existed patient's case, and the database stores the adaption rules. The construction of the database still requires lots of work.

CONCLUSION

The paper proposes a smart assistive system for outpatient care using smart sensing, predictive learning, and intelligent reasoning. The system is proposed to monitor the health condition and provide feedback and assistive guidance to the user. The proposed system has the potential to reduce the cost of health care and provide better services to the patients. Despite the personalized service, the SMASS has the ability to be leveraged to different use cases, such as hospitals and nursing houses. Similar working algorithms can be easily deployed since the infrastructure and software environment of the hospital and nursing houses are similar to the proposed system. The future work could be focused on knowledge database construction algorithms and feasibility analysis.

REFERENCES

AL-Khalidi, F., Saatchi, R., Burke, D., Elphick, H., & Tan, S. (2011). Respiration rate monitoring methods: A review. Pediatric Pulmonology, 46(6), 523–529. https://doi.org/10.1002/ppul.21416

Azad, N., & Lemay, G. (2014). Management of chronic heart failure in the older population. Journal of geriatric cardiology : JGC, 11(4), 329–337. https://doi.org/10.11909/j.issn.1671-5411.2014.04.008

Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. Journal of Personalized Medicine. 2020; 10(2):21. https://doi.org/10.3390/jpm10020021

Bousseljot R, Kreiseler D, Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1 (1995) S 317

Chang, H. H., Hsu, C. C., Chen, C. Y., Lee, W. K., Hsu, H. T., Shyu, K. K., Yeh, J. R., Lin, P. J., & Lee, P. L. (2018). A Method for Respiration Rate Detection in Wrist PPG Signal Using Holo-Hilbert Spectrum. IEEE Sensors Journal, 18(18), 7560–7569. https://doi.org/10.1109/jsen.2018.2855974

Darshan K R and Anandakumar K R, “A comprehensive review on usage of Internet of Things (IoT) in healthcare system,” 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2015, pp. 132-136, doi: 10.1109/ERECT.2015.7499001.

Freyer, J., Greißing, C., Buchal, P., Kabitz, H. J., Kasprick, L., Schuchmann, M., Sultzer, R., Schiek, S., & Bertsche, T. (2016). Entlassungsmedikation – Was weiß der Patient bei Entlassung über seine Arzneimittel? DMW - Deutsche Medizinische Wochenschrift, 141(15), e150–e156. https://doi.org/10.1055/s-0042-108618

H. Gholam Hosseini, K. J. Reynolds and D. Powers, “A multi-stage neural network classifier for ECG events,” 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Istanbul, Turkey, 2001, pp. 1672-1675 vol.2, doi: 10.1109/IEMBS.2001.1020536.

Hospital Statistics by State. (n.d.). American Hospital Directory. Retrieved May 18, 2021, from https://www.ahd.com/state_statistics.html

Huang, M.-J., Chen, M.-Y., & Lee, S.-C. (2007). Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis. Expert Systems with Applications, 32(3), 856–867. https://doi.org/10.1016/j.eswa.2006.01.038

Jha, A. K., Orav, E. J., & Epstein, A. M. (2009). Public Reporting of Discharge Planning and Rates of Readmissions. New England Journal of Medicine, 361(27), 2637–2645. https://doi.org/10.1056/nejmsa0904859

Khera, R., Wang, Y., Nasir, K., Lin, Z., & Krumholz, H. M. (2019). Evaluation of 30-Day Hospital Readmission and Mortality Rates Using Regression-Discontinuity Framework. Journal of the American College of Cardiology, 74(2), 219–234. https://doi.org/10.1016/j.jacc.2019.04.060

L. Cai et al., “OSA patient monitoring system based on the Internet of Things framework,” 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 2019, pp. 1-4, doi: 10.23919/SpliTech.2019.8783103.

L. E. Moser and P. M. Melliar-Smith, “Personal Health Monitoring Using a Smartphone,” 2015 IEEE International Conference on Mobile Services, New York, NY, USA, 2015, pp. 344-351, doi: 10.1109/MobServ.2015.54.

Li, L., Li, Y., Yang, L., Fang, F., Yan, Z., & Sun, Q. (2021). Continuous and Accurate Blood Pressure Monitoring Based on Wearable Optical Fiber Wristband. IEEE Sensors Journal, 21(3), 3049–3057. https://doi.org/10.1109/jsen.2020.3027919

M. Kachuee, S. Fazeli and M. Sarrafzadeh, “ECG Heartbeat Classification: A Deep Transferable Representation,” 2018 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 2018, pp. 443-444, doi: 10.1109/ICHI.2018.00092.

Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

P. L. Penmatsa and D. V. R. K. Reddy, “Smart Detection and Transmission of Abnormalities in ECG via Bluetooth,” 2016 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, USA, 2016, pp. 41-44, doi: 10.1109/SmartCloud.2016.10.

R. Saini, N. Bindal and P. Bansal, “Classification of heart diseases from ECG signals using wavelet transform and kNN classifier,” International Conference on Computing, Communication & Automation, Greater Noida, India, 2015, pp. 1208-1215, doi: 10.1109/CCAA.2015.7148561.

Rachim, V. P., & Chung, W. Y. (2019). Wearable-band type visible-near infrared optical biosensor for non-invasive blood glucose monitoring. Sensors and Actuators B: Chemical, 286, 173–180. https://doi.org/10.1016/j.snb.2019.01.121

S. Saadatnejad, M. Oveisi and M. Hashemi, “LSTM-Based ECG Classification for Continuous Monitoring on Personal Wearable Devices,” in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 515-523, Feb. 2020, doi: 10.1109/JBHI.2019.2911367.

Shahriyar, R., Bari, M. F., Kundu, G., Ahamed, S. I., & Akbar, M. M. (2010). Intelligent Mobile Health Monitoring System (IMHMS). Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 5–12. https://doi.org/10.1007/978-3-642-11745-9_2

Shahriyar, R., Bari, M.F., Kundu, G., Ahamed, S.I., & Akbar, M. (2009). Intelligent Mobile Health Monitoring System (IMHMS). eHealth.

Siddiqui, S. A., Zhang, Y., Lloret, J., Song, H., & Obradovic, Z. (2018). Pain-Free Blood Glucose Monitoring Using Wearable Sensors: Recent Advancements and Future

Prospects. IEEE Reviews in Biomedical Engineering, 11, 21–35. https://doi.org/10.1109/rbme.2018.2822301

DECISION SCIENCES INSTITUTE

Application of Extended Two-Stage Network DEA model to Evaluating and Designing the Humanitarian Supply Chain Network

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

ABSTRACT

In a pre-disaster scenario, the humanitarian supply chain network (HTSCN) design became an important issue long ago. This paper proposes decomposing the single-stage process into the two-stage process and applying the extended two-stage network (ETSN) data envelopment analysis (DEA) method to evaluating and designing the HTSCN for emergency response facilities to function efficiently and effectively. We observe that the proposed approach would provide a more robust and consistent evaluation for designing efficient HTSCN systems through a case study.

KEYWORDS: Humanitarian Supply Chain Network, Extended Two-Stage Network, Data Envelopment Analysis, Multiple Objectives, Emergency Response Facility

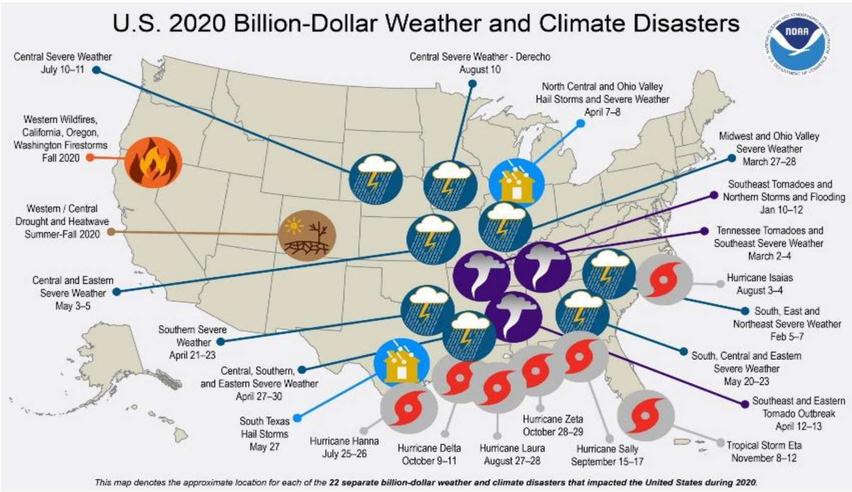
INTRODUCTION

The humanitarian supply chain (HTSC) plays a critical role in providing disaster relief items such as first aids, drinking water, food, and daily commodities to alleviate people's suffering. In early 2019, a wild March was headlined by abnormal warmth in the US's coldest state and a destructive and costly "bomb cyclone" in the central US. By a landslide, Alaska posted its warmest March on record, and the powerhouse storm in the central US became the second billion-dollar weather disaster of 2019. The year 2020 (see Figure 1) sets the new annual record of 22 events, breaking the previous yearly record of 16 events in 2011 and 2017. According to the data developed by the NOAA's National Climatic Data Center (NCDC), the U.S., on average, faces ten severe weather events yearly exceeding one billion dollars in damage. A comparison with an annual average of only two such events throughout the 1980s clearly may force us to speculate that a warming climate could be making these disasters more frequent and more intense.

Many emergencies have brought issues of emergency relief planning again. HTSC is defined as the flow of relief aid and related information between disaster-stricken areas and donors to alleviate vulnerable people's suffering. Indeed, after emergencies, it is critical through emergency response facilities (ERFs) to efficiently and effectively distribute humanitarian aid to the affected areas to save human lives, alleviate suffering, and for a rapid recovery. As mentioned above, logistics planning in emergencies involves the quick and efficient distribution of emergency supplies from the emergency response facilities to the affected areas via supply chains. In this respect, an HTSC design has become an important strategic decision due to the significant damage inflicted by several natural disaster events and the recent COVID-19 pandemic (see Petrudi et al., 2020; van Hoek, 2020). Recently, several authors (Boonmee et al., 2017; Cao et al., 2018; Noham and Tzur, 2018; Hong and Jeong, 2019; Petrudi et al., 2020;

Sarma et al., 2020) consider various HTSC design models; and Zhang et al.(2018) and Liu et al. (2019) have reviewed the papers on the HTSC design problems.

Figure 1: U.S. Billion-dollar weather and climate disasters (2020)



The *ERFs* considered in this paper are three distinctive ones. They are (i) Central Warehouses (*CWHs*) or Distribution Warehouses (*DWHs*), where emergency relief commodities are stored, (ii) intermediate response facilities termed Relief Distribution Center (*RDC*) or Commodity Distribution Point (*CDP*), where people can more effectively gain access to relief goods, and (iii) neighborhood sites (*NBSs*) in need of humanitarian items. An HTSC design problem is inherently strategic and long-term in nature. The main objective of the strategic level is to strengthen emergency preparedness as well as to select the most cost/distance-effective location of *CWHs* and *CDPs* among a set of candidate locations, to establish the distribution of emergency supplies throughout the HTSC, and to assign *NBSs* to *CDPs* and *CDPs* to *CWHs*. In fact, making such a decision is a critical area in designing an effective HTSC system. However, traditional cost-based facility location-allocation models implicitly assume that located facilities will always be in service or be available and do not consider an associated risk of disruption. In reality, all facilities are susceptible to disruptions due to natural disasters, accidents, breakdowns, weather, or strikes. The effects of disruptions could be aggravated due to a lack of flexibility and interdependency in the HTSC. The multi-objective programming (MOP) technique provides an analytical framework where various objectives can be focused on simultaneously so that a decision-maker can use it to provide optimal solutions.

The typical multi-objective programming (MOP) model allows the decision-maker to decide weights for the objective function's deviational variables. It is mainly for reflecting the importance and desirability of deviations from the various goals. However, the actual efficiency of the resulting HTSC is not known. It is unavoidable for the decision-makers to use some of their subjective judgment. How can the best alternative option be selected if the most desirable solution is different among the decision-makers? Hence, it is imperative to answer how to evaluate the efficiency of all alternatives generated by the model and select the most desirable one(s) with a minimized objective function or without any subjective judgment.

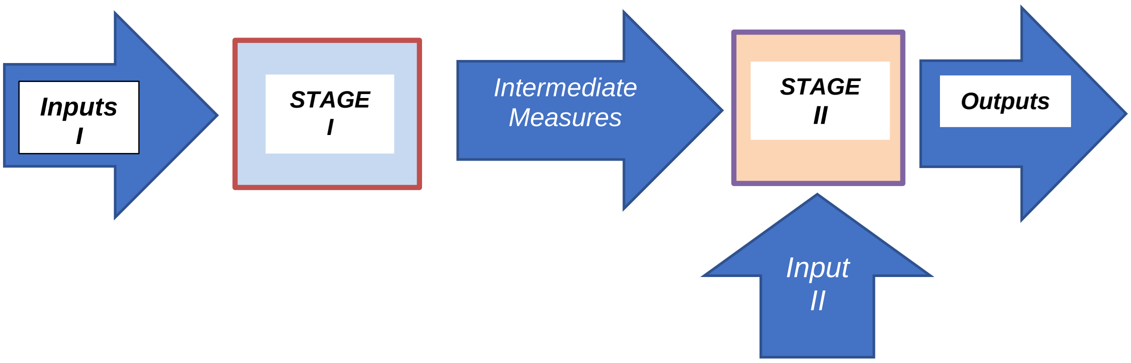
Recently, Hong and Jeong (2019) apply single-stage network data envelopment analysis (SSN-DEA) methods for evaluating and designing the HTSC network (HTSCN) schemes generated by solving multi-objective programming models. The DEA method has been widely accepted as an effective performance evaluation tool. DEA decides the relative efficiency of a set of peer organizations called decision-making units (DMUs). As the whole technique is based on comparing each DMU with all the remaining ones, a large group of DMUs is necessary for the assessment to be meaningful (Ramanathan, 2006). DEA eventually determines which of the DMUs make efficient use of their inputs and produce most outputs and which do not. Each DMU is evaluated with its most favorable weights due to the DEA's nature of the self-evaluation. The DEA model will not consider unfavorable inputs or outputs to raise self-efficiency. As a result, a lack of discrimination is the DEA's major weakness since a considerable number of DMUs out of the set of DMUs to be evaluated are classified as efficient.

To remedy this deficiency, Sexton et al. (1986) propose a cross-evaluation concept to do the peer evaluation rather than the DEA's pure self-evaluation. Doyle and Green (1994) suggest a cross-evaluation matrix for ranking the units by applying the cross-efficiency DEA (CE-DEA) model. Generally, the CE evaluation can provide a full ranking for the DMUs. But, as Doyle and Green (1994) find, the non-uniqueness of CE scores and non-consistent rankings have been critical issues for applying the CE-DEA. But, as the numbers of inputs and outputs of DMUs increase, these issues become more acute in evaluating and ranking them. The super-efficiency DEA (SE-DEA) model concept is introduced to compensate for the SSN-DEA and CE-DEA weaknesses. The resulting model is called a SE-DEA model when a DMU under evaluation is excluded in the DEA models' reference set. The SE-DEA model has significance for discriminating among efficient DMUs. Charnes et al. (1992) use the SE model to study the efficiency classification's sensitivity. Anderson and Peterson (1993) propose the SE model for especially ranking the efficient DMUs.

As Cook and Zhu (2014) mention, an important development area in DEA applications has been that devoted to applications wherein DMUs represent network processes. Thus, in the DEA area literature, a network DEA is one of the significant steams, which controls various sub-stages' efficiencies in a complex structure. In reality, DMUs may consist of two or more stage network structures with intermediate measures. Monfared and Safi (2013) state that the SSN-DEA model considers a DMU as a 'black box' and neglects intervening processes. The SSN-DEA's underlying assumption is that a DMU's performance depends on the inputs used and the output produced. Thus, the 'black box' approaches provide no insights regarding the inter-relationships among the components' inefficiencies and cannot offer specific process guidance to DMU managers to improve DMU's efficiency.

There are two most widely used models for the two-stage network (TSN) DEA method. The first one is that the intermediate measures, which are outputs from the first stage, are the only inputs to the second stage. The second one is called the extended two-stage network DEA (ETSN-DEA) structure, as depicted in Figure 2, where the second stage has two types of inputs; those are its own inputs, and the intermediate measures the outputs from the first stage (see Li et al., 2012). To better evaluate various HTSCN configurations, this paper applies the ETSN-DEA method for the HTSCN design problem in a pre-disaster scenario, consisting of finding the optimal location and allocation of emergency response facilities (ERFs) under the risk of facility disruption. We demonstrate the applicability of the proposed procedure through the case study using the HTSCN model of Hong and Jeong (2019) and compare the results with those generated by the SSN-DEA. The analysis shows that the proposed ETSN-DEA method enables the decision-makers to evaluate the HTSCN schemes more consistently regarding each scheme's efficiency than the traditional SSN-DEA. The proposed approach should be used as an essential tool, along with the conventional DEA, for decision-makers to evaluate and identify the top-rated HTSCN configuration consistently.

Figure 2: Extended two-stage network DEA structure



CASE STUDY AND OBSERVATIONS

Using Hong and Jeong's (2019) study, a case study is considered, using primary disaster declaration records in South Carolina (SC). They cluster forty-six counties based on proximity and populations into twenty counties. Then, one location from each clustered county based on a centroid approach is chosen, assuming that all population within the clustered county exists in that location. The distance between these locations is considered to be the distance between counties. According to the Federal Emergency Management Agency (FEMA) database (FEMA, 2017), SC has experienced sixteen (16) major natural disaster declarations from 1964 to 2017. The database also provides a list of counties where a major disaster was declared. It is assumed that when a major disaster is declared, the county's emergency facility is disrupted and shut down. Based on the historical record, each neighborhood's risk probability (a county or a clustered county) is calculated in Table 2. The potential five locations for CWHs are selected based upon population, the proportion of area that each site would potentially cover, and the proximity to Interstate Highways in SC.

There are 286 configurations for each model arising out of the combinations of the setting of α under the condition. The 286 configurations are reduced to 68 consolidated configurations due to the DMUs with the same four performance measures with different values of the weight set. Each of these 68 network configurations is considered a DMU, representing the optimal locations and allocations of ERFs for a given set of weights. We apply the CRS (Constant Returns to Scale) DEA model for the SSN to find efficient DMUs with a perfect efficiency score (ES) of 1.000. Twenty-six (26) DMUs with CRS ES equal to 1 are identified. The inputs and outputs of these efficient DMUs are decomposed, and the ETSN-DEA is applied for these 26 DMUs. Each DEA method finds a different DMU as the top-ranked DMU. The top-ranked DMUs are selected. The first network configuration finds {Greenville, Charleston} for the CWH locations. The second one finds two CWH locations in the middle of South Carolina, {Columbia, Florence}. The third one finds two CWH locations, {Greenville, Columbia}, far from the coastal area, and looks more balanced. See Hong (2021) for details.

SUMMARY AND CONCLUSIONS

It would be essential to identify the most efficient HTSCN schemes more consistently out of multiple network configurations generated by solving the GP model for various values of the weights assigned to four objectives. This paper transforms the traditional single-stage network (SSN) process into a two-stage process model to apply the extended two-stage network DEA (ETSN-DEA) method.

The results and observations obtained through the case study show that the proposed ETSN-DEA approach works well and should be used as a major method, along with the traditional DEA methods based on the SSN. Identifying several top-rated supply chain network schemes enables the decision-makers to see more options to make the final decisions. Future research would be interesting and necessary to apply the ETSN-DEA extensively in real-world TSN DEA applications.

ACKNOWLEDGMENTS

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, Evans-Allen project number SCX-313-04-18.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE

Logistics Operational Interdependencies for Food Supply Chain Resilience

Muhammad Umar
Lincoln University, New Zealand
Email: muhammad.umar@lincoln.ac.nz

Mark Wilson
Lincoln University, New Zealand
Email: mark.wilson@lincoln.ac.nz

Rizwan Ahmad
Unitec, New Zealand
Email: rahmad@unitec.ac.nz

ABSTRACT

Using a multiple case study approach, this study has investigated operational interdependences of the underlining factors of logistics with food supply chain resilience capabilities in developing economies. Logistics plays a vital role in overall supply chain resilience during the disaster management cycle. In particular, transportation, storage and inventory management are the most critical areas of logistics that need to be fine-tuned in order to quickly and efficiently respond to natural disasters.

KEYWORDS: Supply chain resilience, Logistics management, Disaster management

DECISION SCIENCES INSTITUTE

Exploring Regional Digital Supply Chain Skill Gaps: Are National Trends Confirmed?

Abbas Foroughi
University of Southern Indiana
8600 University Blvd., Evansville IN 47712
Email aforough@usi.edu
Tel.: (812) 465-1667

ABSTRACT

Today's supply chains are undergoing digitalization to increase transparency, stability, and security. Digital technologies often require skills that current employees lack, and many companies are now upskilling and reskilling their workforce. This paper focuses on digital supply chain technology adoption in a particular geographic area, with the purpose of determining challenges and workforce skills needed, the amount of investment in digital technologies, and attitudes toward the role of local universities in providing needed training. The results of this study will facilitate future studies of the extent to which nationally identified trends apply to supply chains in this specific geographic area.

KEYWORDS: Supply Chain Management, Supply Chain Digitalization, Supply Chain Workforce Training, Digital Skill Development, Competency-based Training, Social Learning

INTRODUCTION

From its beginnings as a mere operational function that focused on production lines and delivery to customers, the supply chain has evolved into a management function of its own that focuses on sophisticated planning processes like analytical demand planning, integrated sales, operations planning, and risk management; and is a crucial part of modern digital business (Alicke, Rachor & Seyfert, 2016; Deloitte Development, 2018; Ellis & Santagate, 2018; McKinsey & Company, 2018; Stank, Scott, & Hazen, 2018; Hansen & Malone, 2019; MHI & Deloitte, 2017, 2018, 2019, 2020, 2021). The global pandemic has underscored the crucial role of supply chains in the successful delivery of goods and services around the world. In addition to the economic slowdown caused by covid-19, natural disasters and political upheaval have revealed vulnerabilities that continue to put supply chains at risk (Bergstrom et al., 2020; Dunakin, 2021; Elementum, 2021; BlumeGlobal.com; Isidore, 2021).

Transparency, visibility, control, efficiency, and supplier and customer relationships related to supply chains are even more important in the post-pandemic world (Elementum, 2021; Khan, 2021; SupplyChainGameChanger.com, 2021; Dunakin, 2021). Global supply chains have not yet recovered from the disruptions and bullwhip effect that plagued them during the pandemic (Wolf, 2021), creating even more impetus to adopt digital technologies that provide the transparency and visibility needed to avoid and deal with supply chain uncertainty and risk.

At the same time, the need for supply chain employees with the skills to use them effectively has grown steadily over the past few years (Berry & Mok, 2018; Gonzalez, 2019; Volzer, Burgess & Magda, 2019; Kenco, 2020; Committee for Economic Development, 2020), making upskilling and reskilling current supply chain employees and/or hiring new talent an area of serious concern.

Previous studies have examined digital supply chain technology use at the national and global levels (MHI & Deloitte, 2017, 2018, 2019, 2020, 2021; Alicke, Gupta & Trautwein, 2020). This paper explores the state of the art of digital supply chain technology in a particular region that consists of five Indiana counties centered on the Ohio and Wabash Rivers and Henderson County, in Kentucky, which is directly across the Ohio River from Evansville, Indiana, the largest city in the area, and is closely aligned economically, sharing an Ohio River bridge that facilitates the movement of goods and the workforce back and forth between the two communities. The Evansville region is home to many large, prominent enterprises, in areas such as plastics manufacturing, pharmaceuticals, aluminum, automobiles and automotive parts, electrical power facilities, bicycle manufacturing, mining, concrete and steel production, farming, wood products, agricultural products, custom uniform production, educational institutions, and health care facilities; as well as many small-to-medium businesses and franchisees of national companies such as restaurants, home improvement stores, car dealerships, and retail stores. Evansville is Indiana's third largest city, located 60-miles south of the median center of the U.S. population, resulting in lower transportation costs and easy access to almost every U.S. market (Growth Alliance for Greater Evansville Economic Development). Effective and efficient supply chains play a major role in business and industry in this area.

Insights into reskilling/upskilling needs in this specific area come from studies by Talent 2025, an initiative sponsored by the Evansville Regional Economic Partnership, which encompasses the Economic Development Corporation of SW Indiana, the Southwest Indiana Chamber, and the Growth Alliance for Greater Evansville Economic Development (Talent 2025; Economic Development Coalition of Southwest Indiana). Talent 2025's research has determined that regional manufacturing industry is strong and increasingly automating processes (Whiteside, 2021). However, at the same time, the organization found that current manufacturing employees lack the skills they need to transition to a digital workplace, the future workforce is not being adequately prepared for success in the digital workplace, and many employees risk losing their jobs to changing workplace skill needs (Whiteside, 2021).

With this regional focus in mind, this study will explore the reality of supply chains in the local area, to reveal the adoption/planned adoption of digital supply chain technologies; the need for employee skilling, reskilling, and upskilling; and opportunities for regional colleges and universities to provide customized training that meet the specific needs of regional companies. In particular, the study seeks answers to the following questions:

1. What types of digital technologies are companies using to support their supply chain activities?
2. To what extent do their employees have the skills needed to use digital supply chain technologies?
3. What types/formats of training are companies using to skill, reskill, and upskill supply chain employees to use digital supply chain technologies?
4. How open are companies to facilitating employee training delivered by local universities, and which specific areas of training and training formats would be most valuable to their employees?

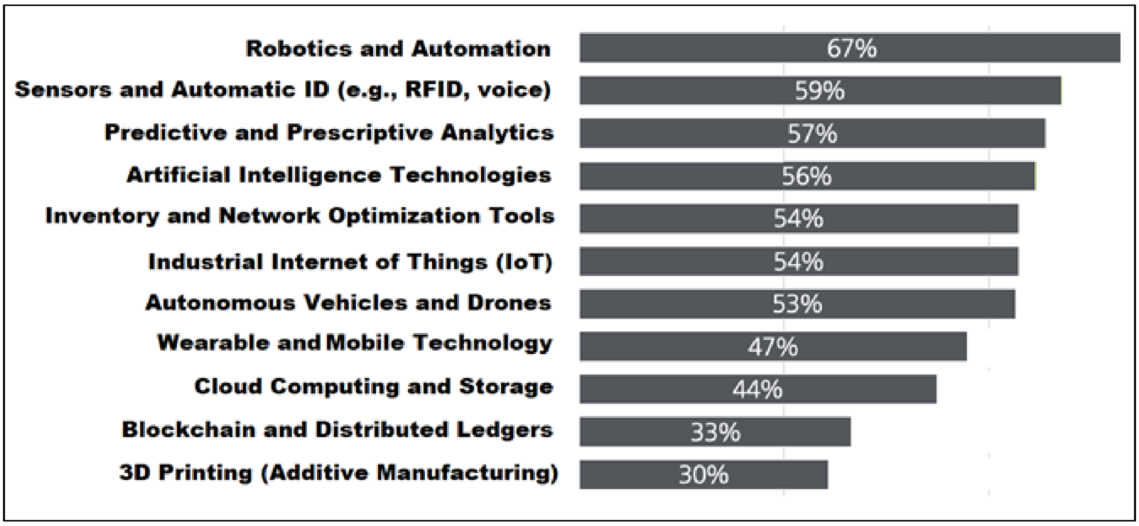
The paper begins with background on digital supply chain technologies and current skill gaps in supply chain employee training, followed by research findings on employee training preferences and sources of supply chain training. The paper finishes with the research methodology used in this study, the findings for each of the above research questions, and discussion and conclusions.

LITERATURE REVIEW

Digital Technologies That Are Transforming Supply Chains

In a survey conducted for the MHI and Deloitte 2020 Annual Industry Report (2020), supply chain professionals identified the following eleven technologies as having potential to disrupt or create competitive advantage for supply chains (See Figure I.)

Figure 1 Technologies with Potential to Disrupt or Create Competitive Advantage



(Adapted from: MHI & Deloitte, 2020, p. 8)

<u>Robotics and Automation</u>

Robotics and automation are used for routine, manual tasks such as picking, packing, inspecting, classifying, and sorting orders; loading, unloading, and stacking; receiving and put-away; assembly operations; and processing such as welding, painting, and cutting (MHI & Deloitte, 2017, 2018, 2019, 2020, 2021; McKinsey & Company, 2018; Treblicock, 2018). Autonomous, self-driving vehicles equipped with sensors can provide data about the environment around them, work together to complete tasks, and provide visibility along the supply chain (Alicke, Rachor & Seyfert, 2016; Stank, Scott & Hazen, 2018; Treblicock, 2018; McKinsey & Company, 2018; MHI & Deloitte, 2018, 2019, 2020, 2021; Banker, 2020). Examples: a parts manufacturer uses robotics to automate picking slow-moving products, saving physical wear-and-tear on employees and speeding up order fulfillment (MHI & Deloitte, 2020). An international energy company adopted Robotic Process Automation (RPA) to improve data collection and reduce requisition-to-pay processes from three days to less than five minutes. An automotive parts supplier partnered with OTTO Motors to install a dual-arm robot onto a self-driving vehicle that receives job requests, navigates to the specified area, and places itself so that the robotic arm can perform the required task. (MHI & Deloitte, 2018).

<u>Predictive Analytics</u>

Predictive analytics uses statistical models, data mining, and machine learning techniques to provide insights about data provided by digital supply chains and to identify the likelihood of future outcomes based on historical data. Predictive analytics analyzes datasets of large size, real-time or near real-time data (Ellis & Santagate, 2018; Benaddi, 2020; RiverLogistics; N-iX, 2021), which are often unstructured and derived from crowdsourcing, internet applications, direct from customers (i.e., point of sale, membership programs, etc.), and other sources. Models capture relationships among many factors to enable the assessment of risk or potential risk associated with a particular set of conditions, and this information is used to forecast consumer behaviors and risk, and to gauge operations efficiency and maintenance needs (MHI & Deloitte, 2017, 2018, 2019, 2020, 2021; Wensing, 2021. Examples: IBM connected data from inventory processes and ERPs to correlate internal and external data in order to achieve data integration and real-time inventory control that increased the accuracy of inventory analyses and predictions (Stank, Scott, & Hazen (2018). Grupo Boticario, one of the largest cosmetics firms, uses IBM Analytics to predict consumer behavior/orders in advance of new product launches or special sales (IBM Business Partners).

The Internet of Things (IoT)

The Internet of Things refers to physical devices, vehicles, home appliances, and other items that are embedded with electronics, software, sensors, actuators, and network connectivity that enable them to connect and exchange data (Deloitte Development, 2018). IoT makes it possible to access or control remotely located objects across the supply chain (Supply Chain Management Review, 2018). IoT enables and facilitates real-time analytics, customer/market insights, customer and/or supplier collaboration, quality control, streamlined production, AI and machine learning, and temperature/humidity monitoring (MHI & Deloitte, 2018, 2019, 2020, 2021; Stank, Scott & Hazen, 2018; Supply Chain Management Review, 2018; Lund et al, 2019; Meola, 2021). Examples: mobile technology, wearable technology, and IoT were used to improve inbound receiving and outbound order picking in a distribution center, as well as product delivery accuracy ((MHI and Deloitte, 2020). A large pharmaceutical company uses CloudLeaf sensors and gateways to track ingredients, providing visibility and monitoring round the clock, as well as a system to access data related to inventory dwelling times and cycle time metrics (MHI and Deloitte, 2021).

Sensors and Automatic ID

Automatic sensory and identification technologies enable robots to sense, identify, and react to, other devices (Buch, 2018; Deloitte Development, 2018; Lund *et al*., 2019; Odum, 2019; Ansari, 2021). Radio Frequency Identification can identify sensory-tagged items up to several meters away and sense thousands of sensory-tagged items at the same time (Treblicock, 2018). Global Positioning Systems (GSPs) receive signals from GPS satellites that enable calculations of an object's position and time, enabling the identification of the exact location of tagged items. Real-Time Locating Systems (RTLS) such as Bluetooth smart phones can identify and track the location of objects within a building or other enclosure, track products and supplies through an assembly line, locate pallets in a warehouse, track temperature and humidity, and send them to the supply chain cloud (Sainathan, 2018; Ansari, 2021). This real-time information is then analyzed by predictive analytics to enable quick responses to anomalies and forecasting to guide future decisions. Example: the University of Washington's Urban Freight Lab designed a Final 50 Feet toolkit that includes the use of sensors and geographic information systems mapping (GIS) to collect data on parking behaviors of commercial vehicles, in order to limit the disruption that occurs in large cities from deliveries (University of Washington).
Traditional identity/governance systems did not have the capacity to provide enterprise-wide visibility on individual user access information, creating the potential of many risk-producing blind spots. Autonomic Identity can now enhance security risk awareness by using AI to collect

and analyze user access points and user account activity across the enterprise, thus identifying risky activity and blind spots (ForgeRock website, 2020; One Network Enterprises, 2021). Example: Benny Auto Group in California adopted a new ID scanner that scans customer's information to the sales and financing departments at the point-of-scan to verify their information. Customers are not allowed to test drive vehicles before their identity is checked and confirmed (eLend Solutions.com).

Inventory and Network Optimization Tools

Inventory optimization tools can provide efficient and effective management of inventory throughout the supply chain, resulting in minimal cost for holding and storage (MHI and Deloitte, 2018). Mathematical algorithms determine the most probable excess stock level and shortage level, helping a company achieve multi-echelon inventory optimization, cost-effective postponement strategies, stock keeping unit rationalization, optimization of inventory components, enhanced supplier intelligence, demand forecasting and planning, and Just-in-Time (JIT) strategies (MHI & Deloitte, 2018; Stank, Scott & Hazen, 2018; DNSStuff.com Staff Contributor, 2021; Miller, Wright, & Brock, 2021). Example: during the Covid-19 pandemic, Ochsner Health, a delivery network of hospitals and clinics in Louisiana, developed an inventory analytics and visualization tool that monitored PPE stock and enabled the company to replenishment stock quicky and support governmental accounting of available PPE supplies (MHI & Deloitte, 2021).
Network optimization tools use mathematical modeling to reduce supply chain network complexities and improve responsiveness to customer needs by optimizing asset locations across the supply chain. Enterprise priorities and supply chain competencies in all areas of the supply chain network (sourcing, transportation, inventory, warehousing) are identified, prioritized, and mapped (Stank, Scott & Hazen, 2018; Lund, Salley & Payne, 2021; Miller, Wright & Brock. 2021). Example: Mars-Rigley, the world's largest confectionary company, brought network optimization to its complex, far-flung operations by using IBM's solution LogicNet Plus to model the current network; create scenarios of procurement, manufacturing, distribution possibilities; identify and test the most favorable scenarios; and select the most robust solutions. The result was a dynamic, real-time model of their supply chain network that answers questions about what, when, where, and how much production is optimal at a given time; and maximizes efficiency and cost-savings by taking into account all of the variables and risks that can occur along their supply chain (Daihes, 2009).

Artificial Intelligence (AI)

Artificial intelligence combines several types of other technologies in the simulation of human intelligence and the rapid solution of complex problems: machine and deep learning, reasoning, voice recognition, augmented reality, cognitive computing, natural language processing, and translation (Ellis & Santagate, 2018; Deloitte Perspectives; Mefford, 2020). AI systems adapt and learn as information, goals, and requirements evolve; and then, they interact easily with other processors, devices, and cloud services (Stank, Scott & Hazen, 2018). AI technologies can improve demand prediction to better match supply with demand at lower cost with less asset investment and better response time, predict maintenance needs, check warehouse manage transportation mode and carrier selection, and manage and/or mitigate risk and disruption self-correction (MHI & Deloitte, 2017; Stank, Scott & Hazen, 2018; Mefford, 2020). Examples: MIT researchers developed an AI model that was trained to discern asymptomatic coughs with 100% accuracy (Entefy blog, Jan 27, 2021). Other examples include robo-bees (bee-sized robots) that have been trained to learn and follow pollination paths using AI and GPS, restoration of prosthetic hand control for amputees with a machine learning algorithm that learns to recognize muscular stimuli that connect to different hand motions, reduction in the amount of time required for new drug development and testing, lasers and machine learning that test the crispness of potato chips, a predictive AI software that can predict where poachers are likely to be, and AUDREY, a virtual agent that enhances fire fighter teamwork by tracking teams and providing updates to firefighters on the team based on their location (Entefy.com blog, Sept. 28, 2020).

Autonomous Vehicles and Drones

With enhanced vehicular sensor technology, driverless cars and trucks are more aware of their surroundings than is possible for human beings. Smart warehousing systems direct trucks to the correct inventory bay, where they are unloaded by robot forklifts, with flying drones inspecting and verifying the load. Examples: Walmart is working with Argo AI and Ford to develop autonomous vehicle delivery service to people's homes (Robotics 24/7, 2021). Automatic guided vehicles (AGV) can now move loads of any size on a factory floor and can be programmed to travel safely throughout the factory (Michel, 2021). EasyJet now uses drone inspections for checking aircraft fuselage for lightning-strike damage, greatly reducing inspection time. DHL uses helicopter drones to deliver packages to hard-to-reach locations such as high elevations and places lacking infrastructure (Hobbs, G. (2018).

Wearable and Mobile Technology

Wearable technologies are application-enabled computing devices, worn or attached to the body, that accept and process input from the Internet or other devices. Workers equipped with wearable technology do not need to input information and can capture, hands-free, leads and sales updates via voice messages. Smart glasses guide staff through the warehouse by virtual reality (Michel, 2017; Supply Chain 24/7, 2014; Van den Bossche, 2016; Herhold, 2020). SmartWatches enable managers to receive 24/7 emails, voice mails, text message, and notifications. Voice control headsets can also provide voice command updates and direct workers to issues or tasks (MHI & Deloitte, 2018). Wearables can improve worker safety by vibrating to warn when an unsafe movement or action has been taken or when a robot or vehicle is approaching (Miller, C., 2021). Examples: Theatro is using artificial intelligence and voice and language recognition systems to interconnect employees with each other and with the enterprise's information, which improves productivity. TaskWatch keeps employees connected with mobile process applications and enterprise systems and customizable notifications. Notable is an AI digital assistant that streamlines and transforms healthcare workflow processes and improves data accuracy and patient interaction (Digitalsilk.com).
Mobile devices now eliminate wasted "travel" time within a warehouse by enabling employees to input/update data and create and print labels right on the spot (Advantech, 2019). That real-time data is then available to everyone in the company, as well as to outside stakeholders, such as supply and distribution partners. With mobile printers, workers can easily print out labels and reports wherever they are, saving time and avoiding costly recording and reporting errors (Advantech, 2019; Redwood Logistics). Example: Fonality's Heads Up DisplayTM enables employees to connect with anyone anytime, from any location or device. On one screen, they can handle phone calls, voicemail, instant messaging (IM), conference calls, email, screen sharing, or video conferences (Fonality.com).

Cloud Computing and Storage

Cloud computing and storage use a network of remote servers to access shared resources like data servers, storage, applications, and other services. Users can store and process data in a privately-owned cloud or a third-party server, making data readily accessible from anywhere. This has the effect of minimizing the cost of infrastructure and maintenance in information technology. Supply chain data stored on a cloud is available to anyone inside the company, as well as distributors and suppliers, so that all of the stakeholders can access the same accurate, real-time information (Alicke, Rachor & Seyfert, 2016; MHI & Deloitte, 2017; Lowe, 2021). Example: When Pfizer transformed its supply chain to “device independence”, by moving to one common cloud-based platform, each supplier was required to implement a cloud computing-based, common-information-exchange framework. The “cloud layer” isolates Pfizer from unexpected changes in its supplier activities and allows supply chain network participants to be added or removed rapidly (Accenture, 2014)

Blockchain and Distributed Ledger Technologies

Blockchain is a continuously growing list of digital records, like a spreadsheet of duplicated records that are linked, secured through cryptography, and continuously updated. Blockchain can be used for transaction processing, records management, and other data-driven tasks. For supply chain clouds, blockchains can provide continuously updated and verified information from data inputs throughout the supply chain in shared ledgers that are available 24/7 to all enterprise stakeholders—internally and externally (Ellis & Santagate, 2018; MHI & Deloitte, 2018; Gstettner, 2019). Examples: A large biotech and pharmaceutical company adopted a Blockchain clinical trial management proof-of-concept solution to provide supply chain visibility and traceability that enabled real-time patient consent collection and sharing and tracking of patient bio-samples during clinical trials, and security for pharmaceuticals throughout the supply chain (MHI & Deloitte, 2017). A shipping company uses blockchain to manage freight tracking that enables buyers, sellers, and officials to track goods shipped around the world. The company collaborated with customs authorities to streamline approvals with a secure blockchain record of transactions and approvals (MHI & Deloitte, 2017). A logistics company has introduced a cloud-based blockchain solution to digitally certify diamonds and protect against unauthorized tampering with supply chain records, certificate reports, and insurance claims. The company uses more than 40 diamond characteristics to create unique diamond IDs. The blockchain solution allows for immutability and security of supply chain data and transparency between diamond certification houses and global diamond suppliers (MHI & Deloitte, 2017).

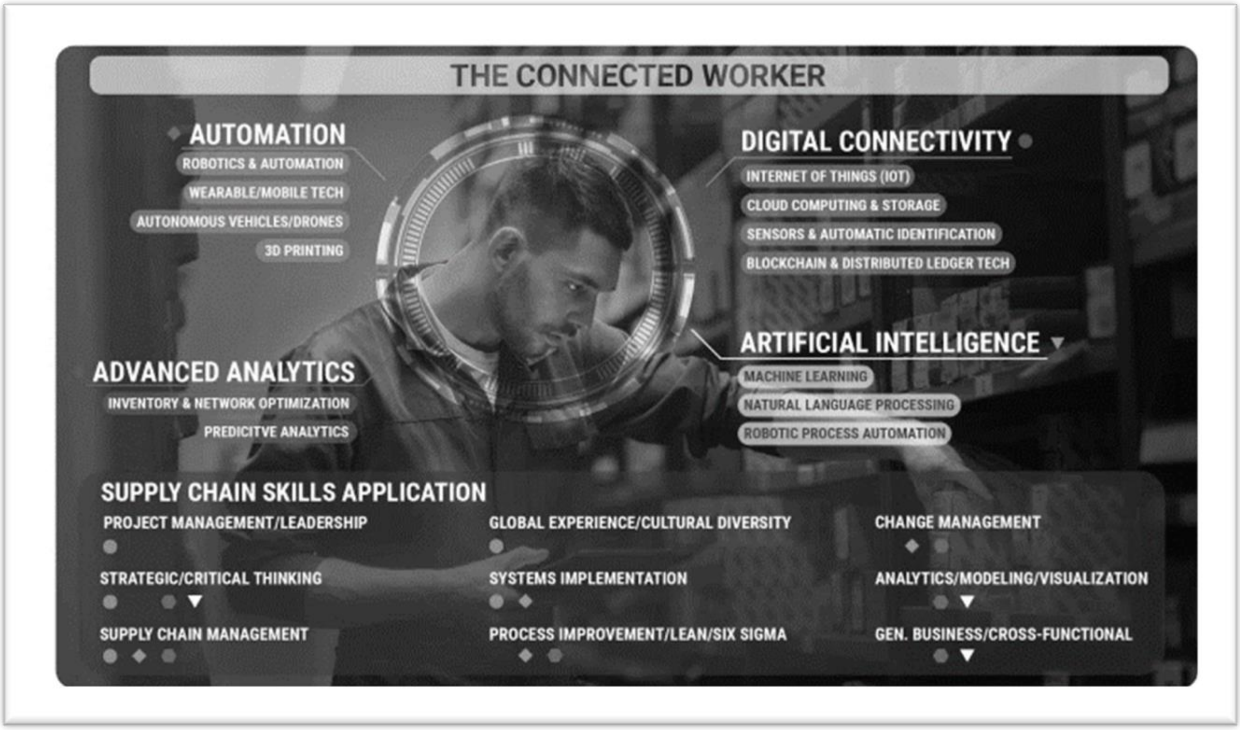
3-D Printing (Additive Manufacturing)

In additive manufacturing, structures are made by the addition of thousands of minuscule layers which combine to create a product. Special CAD software relays messages to a printer, which prints the desired shape in thin layers that are repeatedly printed on top of each other and fused together until the shape is complete (Knowles, 2019). Additive manufacturing reduces material inputs for leaner manufacturing, as well as the cost of production processes, and enables faster reaction to demand changes (Stank, Rachor & Seyfert, 2018; MHI &Deloitte, 2018, 2019, 2020; Knowles, 2019; Moreau, 2021). Examples: For flying ambulances and other airborne vehicles, additive manufacturing can create prototypes of drone components and propeller blades, and quickly copy designs at low cost. Additive manufacturing also supports 5G networks by providing flexibility and agility for the design and production of mass quantities of customized antennae and mounting parts for fast installation (Teipel, 2021). Manufacturers of small-size parts have found benefits from additive manufacturing compared to mold injection for micro parts where orders are small and/or intermittent—more flexibility in case tweaks need to be made, and lower costs for materials, chemicals, molding, and tooling (Digital Engineering 24/7, 2021).

The Critical Need for Supply Chain Employee Training

The covid-19 crisis pointed out the importance of supply chain visibility and flexibility, as well as the important role which digital technologies like the ones described above can play in achieving this (Gonzalez, 2019; MHI & Deloitte, 2021). As digital technologies transform supply chains, employee expertise in mastering and using them is more crucial than ever (Illanes et al., 2018). Upskilling and employee digital dexterity are being increasingly valued more than tenure and experience (McKinsey & Company, 2020). "The Illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn and relearn" (Toffler, 1971, 414).
Figure 2 shows major types of technology which enable today's supply chain employees to be connected—automation, advanced analytics, digital connectivity, and artificial intelligence, and also highlights supply chain skills which employees need to be able to apply to decision making--project management/leadership, strategic/critical thinking, supply chain management,

Figure 2 The Connected Worker (Source: MHI Annual Industry Report, 2020, p. 16)



global experience/cultural diversity, systems implementation, process improvement/Lean/Six Sigma, change management, analytics/modeling/visualization, and general business/cross - functional acumen (MHI & Deloitte, 2020) The *Connected Worker* technology platform from CBT is an example of the way hardware, network, and software tools can be used to enable frontline workers to communicate with supervisors or fellow employees in another location. (CBTechnic.com). Workers in the field can communicate with colleagues or view dashboards through a remote connection. Wearables, with a built-in camera, mic, and speaker with noise-cancellation functionality, are used in addition to laptops, tablets, or phones. A database or resource center facilitates the transmission of real-time actionable information so decisions can be made quickly and accurately. The particular configuration of devices and software can be customized to meet the requirements of different industries. *Connected Worker* can be used in telemedicine to conduct remote examinations and to supplement local resources; and it provides certified, safe equipment and integrated software applications to give oil and gas workers virtual access to sensor readings, P&IDs, maintenance logs, and manuals. IoT sensors enable utility workers to safely install and service equipment, with access to manuals and connection to remote experts. In the manufacturing sector, *Connected Worker* conducts audits and improves maintenance procedures, with step-by-step instructional workflows for new employees and access to connections to remote experts for needed guidance (CBTechnic.com).

The demand for employees with digital skills has already grown by 60% over several years, and this trend is expected to continue or even increase (Berry & Mok, 2018; Agrawal et al., 2020a; Agrawal et al., 2020b; Billing et al., 2021). To emerge stronger from the covid-19 crisis companies need to start reskilling their workforces now. A conscientious habit of continuous, lifetime learning, as well as corporate support for it, will be necessary for employees to succeed in a challenging and evolving work environment like this (Manyuka et al., 2017). Recognizing the lifelong learning message, Walmart has offered to pay 100% of tuition and book cost for any of its associates (Walmart, 2021); and other companies including Amazon, Google, and IBM offer learning/advancement opportunities to employees (Fain, 2019). Alongside critical technical skills, it is also argued that ideally, supply employees should be what Deloitte calls "purple people—employees who have both technical and analytical skills as well as business and communication skills (MHI and Deloitte, 2018). Of course, few individuals possess highly developed sets of both types of skills, but among members of a team, both types of skills should be represented.

Supply chain managers have identified their most pressing workforce challenges (MHI & Deloittte, 2018, 2019, 2020) as hiring (83%), retaining talent (76%), and offering talented staff opportunities for career progression (63%). The MHI & Deloitte 2019 Annual Industry Report emphasized the importance of providing ongoing opportunities in the workplace for employees to develop digital skills, tying career growth to employee participation in training, using younger digital-savvy employees to mentor older employees and assist in facilitating digital transformation, and the establishment of partnerships with STEM programs in universities that encourage the development of future employees with the skills companies need.

The issue of training is particularly critical in the geographic region that is the focus of this study, with research indicating that regional manufacturing employees lack the skills they need to transition to a digital workplace, future employees lack the background on how to succeed in the digital workplace, and current employees risk losing their jobs to changing workplace skill needs (Whiteside, 2021).

Sources of Digital Supply Chain Training

In a previous study, the author identified four sources of digital supply chain training (Foroughi, 2020) that are customizable to the training needs of individual employees or teams. Employees can engage in training that begins where their knowledge and skills are and builds on this to bring them up to speed. All of the providers identified offer courses online, for both individuals and groups, with flexible start/finish parameters, credentials, and customizable formats. First, supply chain training is available from non-profit professional supply chain organizations, such as the Association for Supply Chain Management (ASCM), the Council of Supply Chain Management Professionals (CSCMP), the Institute of Supply Chain Management (IoSCM), and the Institute for Supply Management (ISM). The second category is the consulting company, such as Accenture, the Supply Chain Academy, and Cognitive Class. The third category is MOOC courses (Siemens, 2005), available from Coursera, edX, FutureLearn, iversity, NovoEd, OpenLearn, and Udacity, with a wide variation of training content. The fourth category includes colleges and universities that offer certificates (Gordon, 2018) and MicroMasters programs in supply chain and in a number of technology areas that are important for employees to master for digitalizing supply chains. An example of a successful industry effort to upskill employees with supply chain expertise is General Electric's BrilliantYou collaboration with NovoEd to offer more than 40 courses to its 285,000 employees worldwide through programs at the University of Virginia, the University of California-Berkeley, and Stanford University (NovoEd). The University of Southern California-Marshall's Center for Global Supply Chain Management is collaborating with Infosys, a global leader in consulting, technology and outsourcing solutions, on research and on the development and delivery of supply chain courses (Infosys, 2014). The Palumbo-Donahue School of Business Center for Excellence in Supply Chain Management at Duquesne University and the Association for Supply Chain Management are collaborating to offer industry employees and other adult learners the opportunity to attain three APICS certifications: CPIM (Certified in Planning and Inventory Management), CSCP (Certified Supply Chain Professional), and CLTD (Certified in Logistics, Transportation, and Distribution) (Duquesne University).

Competency-Based Learning and Social Learning

Numerous companies have had success with workplace training programs designed around two pedagogical approaches—competency-based learning and social learning. With the mastery of specific skills in mind, competency-based learning gives learners the flexibility to develop mastery of each competency or skill at their own pace and progress as they demonstrate mastery of academic content, regardless of time, place, or pace of learning (Exner, 2019; Gratton, 2021). Competency-based learning emphasizes the customizability of content, according to the needs of the learner or a group of learners. Learners can develop a specific skill, for which they often receive a badge or other form of validated recognition. Such recognitions are often "stackable", meaning that they can be combined to enable the learner to earn a full qualification, such as a certificate or diploma. Competency-based learning is increasingly delivered fully online, because many students taking such programs are already working or seeking work. Student engagement and outcomes improve, because the content is relevant to each student and tailored to their unique needs (Brown, 2017). Currently, 41.7% of the most profitable companies in the world, in the Fortune 500, including Shell, Toyota, Paypal, and Booking.com, are using online tools for training (Wins, 2019). Such solutions are helpful for training new employees and also help to make ongoing employee education both affordable and effective.

At the same time, social learning theory is also very applicable for workplace learning, because it emphasizes the use of learning from others, mentorships, and online learning in place of formal learning methods like traditional curriculum-based education. With social learning, skill-building opportunities are relevant to specific employee needs, efficient, open, and accessible to all employees, while also offering them greater flexibility and control over their learning. Social learning also emphasizes how employees can learn from each other, either through participation in group learning activities or by observing others. Companies can create and maintain an atmosphere for ongoing learning by rewarding employees who contribute to a culture of learning, creating corporate mentorship programs, and encouraging employees to connect, interact, and learn from each other. With platforms like Skype, Slack, Basecamp, and social media channels, employees can learn from each other by sharing informative content, links, ideas, and other information (Big Think Edge, 2018).

Employee Training Preferences

Coursera for Business (2018) surveyed 750 managers and more than 1,000 employees about on-the-job training, yielding many valuable insights that can guide workplace education efforts, including the following.

- Shared responsibility: Employers and employees agree that they both share responsibility for training—employers for providing it and employees for participating in it.
- Training as a perk: Sixty-three percent of employees said that receiving high-quality training makes them more likely to continue to work for their current employer.
- 62% of employees report being motivated to learn a new skill in order to:
  - improve performance in their current job (42%)
  - build on existing skills in their current job (42%)
  - learn skills for future jobs (36%)
  - increase credibility and expertise (36%)
  - stay up-to-date in industry (31%)
  - feel more confident professionally (28%)
  - learn new skills to complete a task (10%)
  - get a new job (9%).
- Employees prefer content that:
  - is relevant and applicable (49%)
  - is offered in a format that fits their schedule (32%)
  - is valued by colleagues (28%)
  - includes opportunity to earn a valuable credential (23%)
  - is provided by the employer (20%).
- In terms of time commitment for training, employees prefer the following:
  - 1st choice: 45 minutes or less in length
  - 2nd choice: 45-minute to 2-hours
  - 3rd choice: a half-day or longer.
- Employees prefer the following learning formats:
  - Online courses (50%)
  - in-class training (49%)
  - training materials to study on their own (45%)
  - attending conferences (39%) and webinars (39%)
  - consulting friends, co-workers, and mentors) (34%).

The main tenets of competency-based training and social learning-based training—flexibility, online format, customized content, performance-focus, and recognitions—correspond closely to the workplace training preferences identified above in the Coursera survey (2018).

Brief Economic Overview of the Region that is the Subject of this Study

Manufacturing output leads economic activity in the Evansville regional economy, accounting for about 26% of the gross regional product and 13.9% of employment (STATS Indiana). According to the Evansville 2021 Outlook (Khayum, 2020), expansion of economic activity in the Evansville economy was dramatically disrupted in the first quarter of 2020 by the COVID-19 pandemic, although projects and announcements of future fixed investments were expected to lead to increase employment and output in 2021--groundbreaking of a facility for a supplier of food packaging materials, construction of a 100,000-square-foot shell building in an Industrial Park, and an enlarged facility for production of commercial and residential garage door parts and accessories. Economic vitality in the Evansville area is linked to the strength of the broader economy. The income growth gap between the Evansville economy and the U.S. economy over the past three decades underscores the importance of achieving higher rates of future output and employment growth. Upcoming developments in the area include a more than $800 million investment by Toyota Motor Manufacturing-Indiana (Gorman, 2021), $475 million for road projects, and construction of an I-69 bridge connecting Evansville to Henderson, Kentucky (Lyman and Costello, 2021), and a $34.2 million expansion of Berry Global of thermoforming capacity and the addition of new drink cup lines and associated printers (Brown, A., 2021).

RESEARCH METHODOLOGY

This study implemented a survey of the state of the art of digital supply chain technology use in companies in a particular region in southwest Indiana and in a closely-linked city, Henderson, Kentucky. The purpose of the study was first to determine the extent to which digital technologies such as those discussed earlier in this paper are enhancing supply chain management in this specific geographic area, to identify skill gaps companies are experiencing, and to explore opportunities for regional colleges and universities to provide customized training that meets the specific needs of regional companies.

Survey Questionnaire

A survey was developed, based on the survey used by MHI & Deloitte (2020; 2021) and a Coursera for Business study (2018), was emailed via Qualtrix to approximately 900 businesses in the area of interest, with business contacts obtained from the Southwest Indiana Chamber of Commerce and the Henderson, Kentucky Chamber of Commerce. A total of 34 companies responded to the survey.

Results

Survey results were compiled in Qualtrix. The following tables present the survey items, beginning with company demographics questions, followed by the response to the other survey items.

<u>Company Demographics (Tables 1-3)</u>

The role of the individuals who responded to the survey is presented in Table 1.

Table 1: Respondent's Role in the Company		
ROLE	# RESPONSES	% OF TOTAL
CEO (or President)	15	(44.12%)
Other (Specify in box.)	9	(26.47%)
General Manager/Department Head	7	(20.59%)
Vice President or Senior Vice President	3	(8.82%)

Table 2 presents the company descriptions given by the survey respondents.

Table 2: Company Description		
COMPANY DESCRIPTION	# RESPONSES	% OF TOTAL
Other (Specify in box.)	14	(41.18%)
Manufacturing	8	(23.53%)
Service Provider	5	(14.71%)
Consulting	4	(11.75%)
Distributing	2	(5.88%)
Transportation	1	(2.94%)

Healthcare, General Contractor, Construction, Finance-Mortgage Banking, Education-High School, Conventional Supermarket, Development, Retail, Public Media Station, Retail Grocer, Residential and Commercial Solar Installation and Commercial Electric Contractor, Retail Sales, Chamber of Commerce, Advertising

Table 3 shows the size in revenue for the companies that responded to the survey.

Table 3 Company Size in Revenue		
REVENUE SIZE	# RESPONSES	% OF TOTAL
Less than $10 million	22	(64.71%)
$10 million to less than $50 million	6	(17.65%)
Greater than $1 billion	2	(5.88%)
$50 million to less than $100 million	1	(2.94%)
$100 million to less than $500 million	1	(2.94%)
$500 million to less than $1 billion	1	(2.94%)
Do not know	1	(2.94%)

<u>Research Question #1 (Tables 4-7)</u>

What types of digital technologies are companies using to support your supply chain activities? Table 4 presents the types of digital technologies used by companies to support their supply chain activities, now and expected use in and after three years.

Table 4: Technologies used to Support Company's Operations/Supply Chain								
TECHNOLOGY	CURRENTLY USING		EXPECT TO USE IN NEXT 3 YEARS		EXPECT TO USE AFTER 3 YEARS		NOT APPLICABLE	
AI	4	(12.12%)	4	(12.12%)	6	(18.18%)	9	(57.58%)
Automated predictive or preventive maintenance	11	(33.33%)	4	(12.12%)	5	(15.15%)	13	(39.39%)
Automatic vehicles/drones	1	(3.03%)	5	(15.15%)	2	(6.06%)	25	(75.76%)
Blockchain	0	(0.0%)	3	(9.09%)	3	(9.09%)	27	(81.82%)

Cloud computing and storage	30	(88.24%)	2	(5.88%)	1	(2.94%)	1	(2.94%)
Distributed ledger	1	(3.03%)	3	(9.09%)	2	(6.06%)	27	(81.82%)
Industrial Internet of Things (IoT)	3	(9.09%)	7	(21.21%)	1	(3.03%)	22	(66.67%)
Inventory optimization tools	13	(38.24%)	3	(8.82%)	3	(8.82%)	15	(44.12%)
Network optimization tools	17	(50.00%)	3	(8.82%)	1	(2.94%)	13	(38.24%)
Other robots	2	(5.88%)	4	(11.76%)	4	(11.76%)	24	(70.06%)
Sensors or automatic id	12	(35.29%)	5	(14.71%)	1	(2.94%)	16	(47.06%)
Wearable/mobile technology	7	(21.21%)	3	(9.09%)	3	(9.09%)	20	(60.61%)
3D printing	2	(6.06%)	4	(12.12%)	6	(18.18%)	21	(63.64%)

Cloud computing, network optimization tools, inventory optimization tools, sensors or automatic ID, and automated predictive/preventive maintenance are the technologies indicated as the most currently in use. Fifty percent or more Not Applicable ratings were given to blockchain, automatic vehicles and drones, blockchain/distributed ledger, Industrial Internet of Things, automatic vehicles/drones, other robots, Industrial Internet of Things (IIoT), 3D printing, wearable/mobile technology, and AI as being. 3D printing and Other Robots received the highest ratings for Expected Use in/after 3 Years, although these percentages were very modest.

Table 5 presents operational/supply chain challenges faced by responding companies.

Table 5: Operational/Supply Chain Challenges Faced by your Company						
OPERATIONAL CHALLENGES	SELDOM CHALLENGING		SOMETIMES CHALLENGING		ALWAYS CHALLENGING	
Customized products/services	7	(21.21%)	17	(51.52%)	9	(27.27%)
Cyber security risks	5	(14.71%)	17	(50.00%)	12	(35.29%)
Demand forecasting	5	(14.71%)	20	(58.82%)	9	(26.47%)
Faster response times	11	(33.33%)	12	(36.36%)	10	(30.30%)
Food safety, spoilage, and contamination	25	(80.65%)	3	(9.68%)	3	(9.68%)
Hiring/retaining qualified workers	5	(15.15%)	14	(42.42%)	14	(42.42%)
Implementing sustainability	13	(39.39%)	14	(42.42%)	6	(18.18%)

Increasing competitive intensity	9	(27.27%)	11	(33.33%)	13	(39.39%)
Insight into customer behavior	6	(18.18%)	18	(54.55%)	9	(27.27%)
Insight into supply and demand	8	(25.00%)	16	(50.00%)	8	(25.00%)
More frequent shipments	14	(45.16%)	13	(41.94%)	4	(12.90%)
Omnichannel fulfillment/unified commerce	21	(70.00%)	7	(23.33%)	2	(6.67%)
Out-of-stock problems	11	(34.38%)	16	(50.00%)	5	(15.63%)
Product and material level transparency	14	(46.67%)	13	(43.33%)	3	(10.00%)
Rising customer service expectations	7	(21.21%)	14	(42.42%)	12	(36.36%)
Smaller shipments	13	(43.33%)	13	(43.33%)	4	(13.33%)
Supply chain transparency	12	(37.50%)	14	(43.75%)	6	(18.75%)
Synchronization of the supply chain	12	(38.71%)	14	(45.16%)	5	(16.13%)
Visibility of inbound/outbound shipments	15	(48.39%)	14	(45.16%)	2	(6.45%)

Only food safety/spoilage/contamination and omnichannel fulfillment/unified commerce were rated as Seldom Challenges, while all of the other issues were rated Somewhat Challenging or Always Challenging.
Table 6 presents preparations companies are undergoing to prepare for changes in the next ten years.

Table 6: Preparations for Changes in the Next 10 Years

PREPARATION FOR CHANGES IN NEXT 10 YEARS	# RESPONSES	% TOTAL
Partnering with vendors about applications, benefits	24	(27.59%)
Reskilling and training workers for emerging technologies	15	(17.24%)
Increased investment in innovative technologies	14	(16.09%)
Begin piloting new technologies	12	(13.79%)
Recruiting for different skill sets to align with future needs	12	(13.79%)
Changing organizational structure/incentives to create an innovative culture.	9	(10.34%)
Other (Specify in the box.)	1	(1.50%)
Expanding warehouse to stock more product locally.		

Response to Preparation for Changes in the Next 10 Years yielded low percentages, the highest being partnering with vendors about applications/benefits, followed by reskilling and

training workers for emerging technologies, increased investment in innovative technologies, piloting new technologies, and recruiting for different skill sets to align with future needs.
Table 7 presents the amount of investment which surveyed companies are committed to making in supply chain technologies over the next three years.

Table 7: Investment in Supply Chain Technologies over the Next 3 Years		
DOLLAR AMOUNT	# RESPONSES	% TOTAL
Less than $5 million	23	(69.70%)
Do not know	8	(24.20%)
$1 million-less than $5 million	1	(3.03%)
$10 million-less than $50 million	1	(3.03%)
$5 million-less than $10 million	0	(0.00%)
$50 million- less than $100 million	0	(0.00%)
$100 million or more	0	(0.00%)

Almost 70% of surveyed firms plan to spend less than $5 million on supply chain technologies over the next 3 years, with one company planning to spend between $1 million to less than $5, and one from $10-$50 million.

Research Question #2 (Tables 8-9)

To what extent do their employees have the skills needed to use digital supply chain technologies?
Table 8 presents ratings for digital technology skills in the workforce of responding companies.

Table 8: Employee Skills		
SKILLS	# RESPONSES	% TOTAL
Communication and interpersonal skills	24	(17.27%)
General business acumen and cross-functional knowledge	22	(16.00%)
Project management	15	(10.79%)
Data manipulation skills (data modeling, management)	13	(9.35%)
Systems implementation	12	(8.63%)
Analytics, modeling, visualization	10	(7.19%)
Change management	10	(7.19%)
Process improvement/lean/Six Sigma	8	(5.76%)
Visualization/graphic design	8	(5.76%)
Statistics and quantitative skills	7	(5.04%)
Global experience and cultural diversity	5	(3.60%)
Other (Specify in the box)	3	(2.16%)
Supply chain management (e.g., degrees/certifications	2	(1.44%)

Responses to this question were very modest, with only communication and interpersonal skills, general business acumen, cross-functional knowledge, and project management rated above 10%.
Table 9 presents challenges faced by responding companies relating to employees.

Table 9: Challenges Relating to Employees		
CHALLENGES	# RESPONSES	% OF TOTAL
Finding talent	25	(22.73%)
Managing different generational working habits and expectations	17	(15.45%)
Hiring talent	14	(12.73%)
Retaining talent	14	(12.73%)
Reskilling and training existing staff	13	(11.82%)
Measuring/differentiating/rewarding talent	10	(9.09%)
Offering top performers compelling career progression	10	(9.09%)
Willingness of talent to relocate nationally	4	(3.64%)
None	2	(1.82%)
Other (Specify in box.)	1	(0.91%)
Willingness of talent to relocate internationally	0	(0.00%)

Prospective employees' expectation of wages often exceeds what our local retail industry is able to pay; Not all small businesses are owned by multi-million dollar organizations and can't afford to pay $15 per hour (which is an unrealistic minimum wage amount in this area).

Responses were low, but finding, hiring, retaining talent, and reskilling existing employees were the highest.

Research Question #3 (Tables 10-11)

What types/formats of training are companies using to skill, reskill, and upskill supply chain employees to use digital supply chain technologies?
Table 10 presents current training approaches used by responding companies.

Table 10: Current Training Approaches		
TRAINING APPROACHES	# RESPONSES	% TOTAL
Mentoring of new employees by seasoned employees	20	(31.75%)
Building a culture of continual learning	18	(28.57%)
In-house training by HR staff	10	(15.87%)
New learning credentials employees can earn	4	(6.35%)
Virtual/augmented reality to improve training	4	(6.35%)
Mentoring of senior employees by millennial employees	3	(4.76%)
Other (Specify in box.)	3	(4.76%)
None	1	(1.59%)

Supply chain issues do not apply to the majority of our employees.
Third party training, online training.
Currently, only two employees for our business, both with significant supply chain experience and education.

Results indicate the use of building a culture of learning, mentoring of new employees by seasoned employees, and in-house training by HR staff as the most currently used training. Notably, all of these approaches represent in-house efforts and procedures.
Table 11 shows additional sources of supply chain training that are considered by responding companies.

Table 11: Additional Sources of Supply Chain Training Considered		
SOURCES OF SUPPLY CHAIN TRAINING	# RESPONSES	% TOTAL
Online training by non-profit supply chain organizations	15	(39.47%)
Courses/Micro-master degrees offered by universities	10	(26.32%)
Online training offered by supply chain consultant companies	7	(18.42%)
MOOC provider courses	4	(10.53%)
Other (Specify in box)	2	(5.26%)
On-site focused training with credentials. Training by company that sells/installs system.		

Online training by non-profit supply chain organizations, courses/micro-master degrees offered by universities, online training offered by supply chain consultant companies, and MOOC provider courses received the highest ratings.

Research Question #4 (Tables 12-14)

How open are companies to facilitating employee training delivered by local universities, and which specific areas of training and training formats would be most valuable to their employees?
Table 12 shows the perceived helpfulness of customed training offered by local universities.

Table 12: Perceived Helpfulness of Customized Training offered by Local Universities		
HELPFULNESS OF POTENTIAL CUSTOMIZED TRAINING OFFERED BY LOCAL UNIVERSITIES	# RESPONSES	% TOTAL
Somewhat helpful	19	(57.58%)
Very helpful	9	(27.27%)
Not helpful	5	(15.15%)

Nearly 85% of companies surveyed indicated that customized training offered by local universities would be Somewhat/Very Helpful.
Table 13 shows the importance placed by responding companies on specific supply chain training that could be offered by local universities

Table 13: Importance of Potential Supply Chain Training offered by Local Universities						
SUPPLY CHAIN TRAINING	NOT IMPORTANT		SOMEWHAT IMPORTANT		VERY IMPORTANT	
Analytics, modeling, and visualization	2	(6.25%)	17	(53.13%)	13	(40.63%)
Change management	3	(0.38%)	16	(50.00%)	13	(40.63%)

Communication and interpersonal skills	1	(3.03%)	7	(21.21%)	25	(75.76%)
Data manipulation (data modeling and data management)	3	(9.09%)	15	(45.45%)	15	(45.45%)
General business acumen and cross-functional knowledge	1	(3.03%)	8	(24.24%)	24	(72.73%)
Global experience and cultural diversity	8	(25.00%)	16	(50.00%)	8	(25.00%)
Process improvement, Lean/Six Sigma	3	(9.09%)	15	(45.45%)	15	(45.45%)
Project management	3	(9.09%)	9	(27.27%)	21	(63.64%)
Statistics/quantitative skills	3	(9.09%)	15	(45.45%)	15	(45.45%)
Strategic/problem-solving/critical thinking	1	(3.03%)	6	(18.18%)	26	(78.79%)
Supply chain management (e.g., degrees/certification)	4	(12.12%)	19	(57.58%)	10	(30.30%)
Systems implementation	3	(9.09%)	18	(54.55%)	12	(36.36%)
Visualization/graphic design	8	(25.00%)	13	(40.63%)	11	(34.38%)

All training areas except global experience/cultural diversity and visualization/graphic design were rated as 75% Somewhat/Very Important, with the majority of topics rated as 90% or more. Table 14 shows training formats preferred by responding companies.

Table 14: Preferred Training Course Formats		
PREFERRED TRAINING FORMATS	# RESPONSES	% TOTAL
Online	26	(17.45%)
45-minute to 2-hour training	19	(12.75%)
Attending conferences and webinars	18	(12.08%)
Flexible start-finish	15	(10.07%)
In-class training	11	(7.38%)
Consulting friends, co-workers, and mentors	10	(6.17%)
45 minutes or less in length	9	(6.04%)
Customized content	9	(6.04%)
Opportunity to earn a valued recognition/certificate/badge	9	(6.04%)
Half-day or longer	5	(3.36%)
Performance-focus	4	(2.68%)
Relevant books, blogs, or magazines	4	(2.68%)
None	1	(0.67%)

Although ratings are generally low (17.45% or less), online format, followed by 45-minute to 2-hour training, attending conferences and webinars, and flexible start-finish received the highest ratings, indicating preference for short, online, flexible conferences and webinars.

DISCUSSION AND CONCLUSIONS

This study examined business/industry in a particular geographic area, focusing on the extent to which the region's businesses/industries have adopted digital technologies to support supply chains, and their particular employee skill and training needs, The ultimate purpose was to reveal skill and training skill gaps that can be filled by programs offered at local universities. The following discussion includes comparisons to results from a national/international study (MHI & Deloitte, 2021) and to a Coursera for Business survey (2018).

Research Question #1

Technologies in Use Now or Expected to be Use in/after 3 Years

Results for this item showed low to modest use of the technologies included in the survey, with the exception of cloud computing, which is used by 88.24% of companies surveyed and network optimization tools, used by 50% of companies. Only three other technologies were used by more than 30% of companies--automated predictive or preventive maintenance, inventory optimization tools, and sensors or automatics ID. The remaining of the 13 technologies were rated as Not Applicable. Of the top 10 technologies identified in the MHI & Deloitte (2021), only cloud computing and storage, inventory and network optimization tools were rated highly. These results from Table 4 become more interesting when compared with other results to follow in this discussion.

Operational Challenges

Results for this item revealed striking similarities between the local company results and those of national/international companies surveyed by MHI & Deloitte (2021). With the exception of smaller shipments, all of the top 10 issues in the local study are the same as those in the national/international study, although in somewhat different order. The supply chain challenges identified in this survey item are interesting and predictable in light of the overall modest adoption of technologies that could help to address and overcome them.

Preparation for Changes in the Next 10 Years

Again, responses were low, indicating that local companies for the most part are not proactively seeking ways to address changes that are expected in the next 10 years. However, the top-rated mode of preparation identified by the local companies, partnering with vendors about applications, benefits, is also at the top of the list for national/international companies. Companies appear to want to first rely on vendors to support them in selecting and implementing new technology, but realize the importance of reskilling/training workers, increased investment in technologies, and recruiting for different skill sets.

Amount to be Invested in Supply Chain Technologies over the Next Three Years

Results indicated that the vast majority of local companies will invest less than $5 on supply chain technologies over the next three years. This is consistent with the results from Table 6, which indicated low amounts of preparation for changes in the next ten years.

Research Question #2

Employee Skills for Using Digital Technologies

Results indicated that companies feel that employee skills are lacking in many areas. The highest ratings were given to communication and interpersonal skills, general business acumen/cross-functional knowledge, and project management. Interestingly, the top three skills rated as lacking are what are considered to be "soft skills," not technical skills. This result aligns with what has been described as the need for "purple people" who have both soft skills and are digitally-savvy (MHI & Deloitte, 2018).

Challenges Relating to Employees

Finding, hiring, and retaining talent and reskilling and training existing staff, along with managing different generational work habits and expectations, are the top five challenges related to talent for local companies. These results are similar to the results for national/international companies, who rated finding, hiring, and retaining talent in the top five challenges, with reskilling and training existing staff as the sixth top challenge (MHI & Deloitte, 2021).

Research Question #3

Training Approaches Currently Used

Results indicated the predominant use of mentoring of new employees by seasoned employees, building a culture of continual learning, and in-house training by HR staff. Interestingly, all three of these approaches are offered in-house.

Additional Sources of Supply Chain Training Considered

Results indicated preferences for online training by non-profit supply chain organizations and courses/micro-master degrees offered by universities, indicating a recognition of the need to reach outside the organization for training options.

Research Question #4

The results for this research question are very positive and indicate a strong interest in the role of local universities in building the digital skills needed for supply chain employees.

Perceived Helpfulness of Customized Training Offered by Local Universities

Results showed positive attitudes of more than 85% toward this type of training, which is encouraging for local university efforts to offer customized training to employees.

Importance of Specific Supply Chain Topics

Results indicated that all topics except global experience/cultural diversity and visualization/graphic design are considered from 75%-to more than 90% Somewhat Important to Very Important.

Preferred Training Formats

Responses were low to modest, with the strongest support given to online format, followed by 45-minute to 2-hour training, conferences and webinars, and flexible start and finish. This information will be valuable in designing employee training courses.
To summarize, this survey study has shown that local companies are experiencing challenges related to both operations/supply chain and lack of employee skills needed for digital supply

chain success. Most training is being conducted internally. However, results also indicate positive attitudes toward the possibility of employee training outside the company, as well as toward the idea of training offered by local universities. The results of the study will be shared with local companies in the area that is the focus of this study, with the University of Southern Indiana, the Evansville Regional Economic Partnership, and the Southwest Indiana Chamber of Commerce, as well as the Henderson, Kentucky Chamber of Commerce, so that collaborations can be forged to develop training in the specific supply chain skill areas which local companies need.

The importance of this type of business outreach is reflected in the USI Strategic Priorities, about which President Ronald Rochon said, "…we must be a catalyst for change and a university on the front end of creating a talented, educated citizenry that meets the entrepreneurial and workforce needs of our society" (University of Southern Indiana Strategic Plan) Furthermore, Strategic Priority 3 of USI's Romain College of Business expresses the college's commitment to "deliver visible impacts on society at the local, regional, state and national level," and the corresponding goals, which include the goals of [providing] collaborative opportunities with employers, alumni and other RCOB stakeholders for both faculty and students" and [providing] thought leadership through research and engagement activities" (University of Southern Indiana Strategic Plan).

The response to the survey was admittedly not strong, thus limiting the validity of the results obtained. However, the results did accomplish its stated purpose of shedding light on local supply chain skills needs, and the strong support shown for training support offered by local universities is encouraging and significant.

A strength of the study is the fact that local areas are not often the subject of survey studies, so that study results are not always applicable to local situations. Surveys like the current one reveal realities in specific areas and point out areas of potential improvement and opportunity. Sharing the results with local business and industry will highlight digital technology skill needs, raise the awareness of other companies about this issue, and potentially increase the number of companies that seek training from local universities. It will also contribute to the effort by organizations like the Evansville Regional Economic Partnership to boost economic activity in the region through workforce development that prepares employees for success as a digital-trained workforce (Whiteside, 2021). The collaboration which will potentially result from this study for customized training between local companies and universities, the types of training that are most successful, and the attitudes of employees toward this training will be interesting areas for further research investigation.

REFERENCES

Accenture (2014). Supply chain management in the cloud: How can cloud-based computing make supply chains more competitive? Retrieved from Accenture.com on Sept. 2. https://www.accenture.com/_acnmedia/accenture/conversion-assets/dotcom/documents/global/pdf/dualpub_1/accenture-supply-chain-management-in-the-cloud.pdf

Advantech (2019). How to select mobile pcs in an expanding supply chain technology climate. Advantech White Paper, January 4, 2019. Retrieved from adventech.com, April 25, 2021.

Agrawal, S., De Smet, A., Lacroix, S. & Reich, A. (2020a). To emerge stronger from the covid-19 crisis companies should start reskilling their workforces now. McKinsey & Company Organization Practice & McKinsey Accelerate. Retrieved from mckinsey.com, May 9, 2021.

Agrawal, S., De Smet, A., Poplawski, P. & Reich, A. (2020b). Beyond hiring: how companies are reskilling to address talent gaps. McKinsey & Company. Retrieved from mckinsey.com, April 25, 2021.

Alicke, K., Rachor, J. & Seyfert, A. (2016). Supply chain 4.0—The next-generation digital supply chain. McKinsey & Company Operations Article. Retrieved from mckinsey.com, April 19, 2021.

Alicke, K., Gupta, R. & Trautwein, V. (2020). Resetting supply chains for the next normal. Mckinsey Survey, October 20. Retrieved from mckinsey.com, March 28, 2021.

Ansari, M.A. (2021). The yings and yangs of automation in supply chain. *Journal of Internet Banking and Commerce* 26 (1) January.

Banker, S. (2020). Robots and the autonomous supply chain. Forbes.com, April 2. Retrieved from forbes.com, April 21, 2021.

Benaddi, P. (2020). Impact of predictive analytics on supply chain. Retrieved from daacap.com, March 27, 2021.

Bergstrom, J.E., Gallagher, P.R. & Stewart, I.A. (2020). Looking beyond the horizon: preparing today's supply chains to thrive in uncertainty. *Deloitte Insights*. Deloitte Center for Integrated Research, Deloitte Development, LLC, Dec. 11. Retrieved from www2.deloitte.com, March 20, 2021.

Berry, D. & Mok, L. (2018). Develop the competencies your workforce needs for the digital ecosystem. *Gartner.com*, Refreshed: 05 December 2018 | Published: 21 July 2017. Retrieved from gartner.com, March 24, 2021.

Big Think Edge (2018). How do you apply social learning theory in the workplace?,July 31, 2018. Retrieved from bigthinkedge.com, March 15, 2021.

Billing, F., DeSmet, A., Reich, A. & Schaninger, B (2021). Building workforce skills at scale to thrive during—and—after the covid-19 crisis. *McKinsey & Company Survey*. April 30.

Blume Global.com. What are the main supply chain challenges? Retrieved from Blume Global.com, April 1, 2021.

Brown, A. (2021) Berry global planning $110M expansion. *Inside Indiana Business*, Sept. 14, 2021, 4:32 PM EDT.

Brown, B. (2017). Social learning theory in the workplace. *Bizfluent.com*. September 26. Retrieved from bizfluent.com, April 4, 2021.

Buch, N. (2018). Technology in supply chain management: an overview. *ScaleFusion* Blog, June 29. Retrieved from scalefusion.com, April 3, 2021.

Committee for Economic Development (CED) (2020). The future of work: how america can meet the upskilling challenge. *2020 Solutions Brief.* Retrieved from www.ced.org, April 26, 2021.

CBTechnic.com. Enabling workers to collaborate safely & efficiently from remote locations Retrieved on Sept.3, 2021 from https://www.cbtechinc.com/connected-worker/

Coursera for Business. (2018). Learning & development from both sides of the table: 21 insights to maximize your l & d investment. Retrieved from www.coursera.com, April 30, 2021.

DNSStuff.com Staff Contributor (2021). 6 best network inventory tools and software in 2021. *DNSStuff.com*, December 2. Retrieved from dns stuff.com, April 22, 2021.

Daihes, K. (2009) Global supply chain network design. Presented at the IBM Smarter Supply Chains Regional Symposium and User Conference, Chicago, IL. https://www.youtube.com/watch?v=lonezN8mJSQ

Deloitte Development (2018). Using smart sensors to drive supply chain innovation. Retrieved from www2.deloite.com, April 30, 2021.

Deloitte Perspectives. Cognitive supply chain and a new way of working your data and infrastructure with artificial intelligence. Retrieved from www2.deloitte.com, April 27, 2021.

Digital Engineering 24/7.com. (2021). Making the case for micro-precision 3d printing. https://www.digitalengineering247.com/download/making-the-case-for-micro-precision-3d-printing/products. August 30, 2021.

DigitalSilk.com. How wearable technology In business drives innovation forward. Retrieved from https://www.digitalsilk.com/wearable-technology-in-business, Sept. 4, 2021.

Dunakin, C. (2021). 28 supply chain professionals share the biggest challenges of supply chain management, 6 River Systems blog., Feb. 10. Retrieved from riversystems.com, March 17, 2021.

Duquesne University Palumbo-Donahue School of Business Center for Excellence in Supply Chain Management. New Master's Degree in Supply Chain Management and Certificate. Retrieved from Duquesne University - School of Business , Sept 20, 2021.

Economic Development Coalition of Southwest Indiana. Comprehensive economic development strategy. Retrieved from southwest Indiana.org/ceds, May 6, 2021.

Elementum (2021). Three fears looming over supply chains in 2021. Whitepaper. Retrieved from elementum.com, February 7, 2021.

eLend Solutions Case Study. New id scanner eliminates id frau at dealerships while also boosting sales. Retrieved from Lend Solutions.com on August 12, 2021. (https://www.elendsolutions.com/press_release/case-study-new-id-scanner-eliminates-id-fraud-at-dealerships-while-also-boosting-sales-2/

Ellis, S. & Santagate, J. (2018). The path to a thinking supply chain. *IDC Custom Solutions Technology Spotlight*, August.

Entefy blog (Sept. 28, 2020). From robot bees to crunchier potato chips: 6 creative uses of ai today. Retrieved from Entefy.com, Sept. 1, 2021.https://www.entefy.com/blog/post/605/from-robot-bees-to-crunchier-potato-ch,ips-6-creative-uses-of-ai-today

Entefy blog (Jan. 27, 2020). The intelligent enterprise leaps forward in 2021, Retrieved from The intelligent enterprise leaps forward in 2021 , Sept. 1, 2021.

Exner, P. (2019). Why is competency-based training so effective? SkillBuilderLMS, May 1 2019. Retrieved from skillbuilderslms.com, March 29, 2021.

Fain, P. (2019). Employers as educators. *Inside Higher Education*. July 17, 2019. https://www.insidehighered.com/digital-learning/article/2019/07/17/amazon-google-and-other-tech-companies-expand-their

Fonality. Our unified communications solution heads up display. Retrieved from fonality.com, Sept. 5, 2021.
https://www.fonality.com.au/solutions/heads-up-display?__hstc=43548986.c6c45af10a71b91a77b3210ee6902890.1632245539982.1632245539982.1632245539982.1&__hssc=43548986.2.1632245539982&__hsfp=3520052351&_ga=2.161578492.1153095702.1632245537-904464590.1632245537

ForgeRock.com. Autonomous identity product fFactsheet. Retrieved from forgerock.com, April 18, 2021.

Foroughi, A. (2020). Supply chain workforce training: addressing the digital skills gap. *Higher Education, Skills, and Work-based Learning*, September 18, 2020. https://doi.org/10.1108/HESWBL-07-2020-0159

Gonzalez, A. (2019). The most important skills for supply chain young professionals. *Talking Logistics*. *Adelante SCM*, July 10, 2019. Retrieved from talkinglogistics.com, March 29, 2021.

Gordon, A. (2018). 'MicroMasters' surge as moocs go from education to qualification. *Forbes*.com, Feb. 13, 2018.

Gorman, Evan (2021). Toyota’s expansion to bring significant economic impact to greater Evansville area. Evansville Courier Online, April 28, 2021, 10:41 PM CDT.

Gratton, L. (2021). The Future of workplace learning: To create a learning infrastructure that enables a more equitable workforce, corporations must take bold steps. *MIT Sloan Management Review*, March 8, 2021

Growth Alliance for Greater Evansville Economic Development. Retrieved from growthallianceevv.com, May 4, 2021.

Gstettner, S. (2019). How blockchain will redefine supply chain management. *Operations Management Podcasts*, Knowledge@Wharton, July 30. Retrieved from knowledge.wharton.upenn.edu, May 2, 2021.

Hansen, D. & Malone, T.W. (2019). The dawn of the intelligent enterprise: artificial intelligence and machine learning power the new workforce. *MIT Sloan Management Review*, Sponsor's Content (Oracle), January 7, 2019.

Herhold, K. (2020). Three ways wearables boost supply chain efficiency. *SupplyChainBrain* Think Tank Blog, January 31. Retrieved from supplychainbrain.com, May 5, 2021.

Hobbs, G. (Dec. 19. 2018) Case study: top 5 companies that are using drones. *Commercial UAV News*. https://www.commercialuavnews.com/construction/case-study-top-5-companies-that-are-using-drones

IBM Business Partners. Grupo Boticario: Predicting consumer demand for cosmetics with might that is more than skin deep. Retrieved from https://mediacenter.ibm.com/media/Grupo%20Botic%C3%A1rio%20%20-%20Predicting%20consumer%20demand%20for%20cosmetics%20with%20insight%20that%20is%20more%20than%20skin-deep/1_anhqqzsg, Sept. 15, 2021.

Illanes, P., Lund, S., Mourshed, M., Rutherford, S. & Tyreman, M. (2018). Retraining and reskilling workers in the age of automation. *McKinsey*, January 22, 2018. Retrieved from www.mckinsey.com, April 27, 2021.

Infosys (2014) Infosys and usc-marshall school of business announce industry-academia partnership for global supply chain management initiative to help address demand for over one million supply chain professionals in the us by 2016. February 5, 2014. Infosys Newsroom https://www.infosys.com/newsroom/press-releases/2014/industry-academia-partnership-gscm.html

Isidore, C. (2021) Automakers' problems are much worse than we thought, *CNN Business,* 1:45 PM ET, Tues., September 28, 2021. https://www.cnn.com/2021/09/28/business/auto-industry-supply-chain-problems/index.html

Kenco (2020). State of the supply chain: 2020 innovation survey. Kenco Whitepaper. Retrieved from kencogroup.com, May 4, 2021.

Khan, O. (2021). Digitalized supply chains: building the future one step at a time. *SkillDynamics* Whitepaper. February. Retrieved from skilldynamics.com, May 8, 2021.

Khayum, M. Evansville forecast, 2021. (2020). *Indiana Business Review*, 2020.

Knowles, S. (2019). Five ways 3d printing will impact the global supply chain. *Main Pointe Practical Insight* blog, March 6. Retrieved from mainpointe.com, March 28, 2021.

Lowe, H. (2021). How cloud supply chain software revolutionizes operations. *SelectHub*, 2021. Retrieved from selecthub.com, April 4, 2021.

Lund, P, Salley, A. & Payne, T. (2021). Magic quadrant for supply chain planning solutions. *Gartner.com*, February 22. Retrieved from gartner.com, May 1, 2021.

Lund, S., Manyika, J., Woetzel, J., Bughin, J. Krishnan, M., Seong, J. & Muir, Mac. (2019). Globalization in transition: The future of trade and value chains. McKinsey Global Institute. Retrieved from mckinsey.com, May 2, 2021.

Lyman, J. and Costello, J. (2021) Gov. Holcomb announces $475 million in road work, including new highway. 14news.com, June 11, 2021, 6:13 AM CDT. https://www.14news.com/2021/06/11/gov-holcomb-visiting-evansville-friday/

MHI & Deloitte. (2017). The 2017 mhi annual industry report: next-generation supply chains: digital, on-demand and always-on. Retrieved from mhi.org, March 25, 2021.

MHI & Deloitte (2018). The 2018 mhi annual industry report: overcoming barriers to next-gen supply chain innovation. Retrieved from mhi.org, March 2, 2021.

MHI & Deloitte (2019). The 2019 mhi annual industry report: elevating supply chain digital consciousness. Retrieved from www.mhi.org, March 22, 2021.

MHI & Deloitte (2020). The 2020 mhi annual industry report: embracing the digital mindset—connecting data, talent, and technology in digital supply chains. Retrieved from mhi.org, March 2, 2021.

MHI & Deloitte (2021). The 2021 mhi annual industry report: innovation driven resilience—how technology and innovation help supply chains thrive in unprecedented times. Retrieved from mhi.org, April 2, 2021.

Manyuka, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R. & Sanghvi, S. (2017). Jobs lost, jobs gained: what the future of work will mean for jobs, skills, and wages. Retrieved from mckinsey.com, April 4, 2021.

McKinsey & Company (2018). The automation imperative survey, September. Retrieved from mhi.org, April 24, 2021.

McKinsey & Company (2020). The next normal: the future of capability building. Retrieved from mckinsey.com, April 6, 2021.

Mefford, D. (2020). How AI Is Transforming Global Supply Chains. *SupplyChainBrain* Think Tank, December 28. Retrieved from supplychainbrain.com/blogs, April 29, 2021.

Meola, A. (2021). How ai and iot devices will revolutionize supply chain logistics and management in 2021, *Insider*, Feb. 22. Retrieved from businessinsider.com, May 5, 2021.

Michel, R. (April 28, 2017). Information management: wearables come in for a refit. *Supply Chain 24/7.com*. Retrieved from supplychain24/7, May 30, 2021.

Michel, R. (August 24, 2021). ACVs gain in flexibility and applications. *Modern Materials Handling*. https://www.mmh.com/article/agvs_gain_in_flexibility_and_applications

Miller, C. (2021). Beyond resilience: the role of robotics and advanced technologies in making facilities safer. *MHI Solutions*, 10(9), Sept. 22, 2021.MHI Solutions (MHIQ) - Volume 9, Issue 4 - page 10 (mhisolutions-digital.com)

Miller, M., Wright, J. & Brock, D. (2021). Realising end-to—end transparency in the supply chain. *Supply Chain Digital.com*, May, pp. 37-45. Retrieved from supplychaindigital.com, May 2, 2021.

Moreau, C. (2021). The state of 3d printing report: 2021. Survey, *Sculpteo eBooks*. Retrieved from sculpteo.com, April 21, 2021.

N-iX. Big data and predictive analytics in supply chain: success stories and tips. Retrieved from n-ix.com, April 26, 2021.

NovoEd. From crotonville to digital: upskilling ge's global workforce. Retrieved from https://www.novoed.com/wp-content/uploads/2020/05/GE_CaseStudy.pdf, August 29, 2021.

Odum, S.M. (2019). Tech briefing: auto id. *Supply Management*, April 5. Retrieved from www.cips.org, April 11, 2021.

One Network Enterprises (2021). Top trends accelerating supply chain operations in 2021. *Supply Chain 24/7*, Nov. 24. Retrieved from supplychain247.com, March 16, 2021.

Redwood Logistics. What is a Mobile Supply Chain? Retrieved from redwoodlogistics, April 3, 2021.

RiverLogistics.com. Supply chain predictive analytics: what is it and who's doing it. Retrieved from riverlogic.com, May 5, 2021.

Robotics 24/7 (Sept. 16, 2021). Argo AI, Ford, and Walmart to launch autonomous vehicle deliveries in 3 cities. Argo AI, Ford and Walmart to Launch Autonomous Vehicle Delivery Service in Three U.S. Cities - Argo AI .

Sainathan, P. (2018). 6 Ways bluetooth beacons can improve inventory and warehouse management. *Roambee blog*, November 19, 2018. Retrieved from blog.roambee.com, May 2, 2021.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2, 3-10.

Stank, T., Scott, S. & Hazen, B. (2018). A savvy guide to the digital supply chain: how to evaluate and leverage technology to build a supply chain for the digital age. University of Tennessee Global Supply Chain Institute's Technology in the Supply Chain Series.

STATS Indiana, using U.S. Bureau of Labor Statistics and Indiana Department of Workforce Development data. Indiana (May 2020) OEWS State Occupational Employment and Wage Estimates (bls.gov).

SupplyChainGameChanger.com. (2021). Preparing business for the digital supply chain: oracle interviews mike mortson!. Supply Chain Game Changer. March 15. 2021. Retrieved from supplychaingamechanger.com, May 3, 2021.

Supply Chain 24/7. (2014). Next-Generation warehouse control system on the wearable smartwatch, March 16, 2014. Retrieved from supplychain247.com, April 16, 2021.

Supply Chain Management Review (2018). How iot enables an intelligent and connected supply chain. SCMR White Paper, September 28. Retrieved from www.scmr.com, April 24, 2021.

Talent 2025: Greater evansville talent initiative. Retrieved from growthallianceevv.com/blog/talent-2025-greater-evansville-talent-initiative, March 27, 2021.

Teipel, B. (2021) Esentium.com. How 3d printing can save the world…and maybe your life.Retrieved from https://www.essentium.com/how-3d-printing-can-save-the-worldand-maybe-your-life/ , August 24, 2021.

Toffler, A. (1971). *Future Shock*. Bantam Books: New York City, New York.

Treblicock, B. (2018). The warehouse of the future. *Logistics Management.com*, December 20, 2018. Arizona State University White Paper, W. P. Carey School of Business.

Van den Bossche, P. (2016). Wearable technology in the warehouse. Supply Chain 24-7, February 1, 2016. Retrieved from supplychain247, April 16, 2021.

University of Southern Indiana Strategic Plan. Retrieved from https://www.usi.edu/strategic-plan/, Sept. 4, 2021.

University of Washington. Urban Freight Lab. The final 50 feet: urban goods delivery system toolkit. Retrieved from https://depts.washington.edu/toolkit/home#:~:text=The%20Urban%20Freight%20Lab%20at,dock%20or%20internal%20freight%20bay, Sept. 15, 2021.

Volzer, D, Burgess, J. & Magda A. (2019). Reimagining the workforce 2021: closing the skills gap through education. Wiley Education Services, & Future Workplace. Louisville, KY: Wiley edu, LLC.

Walmart (2021). Walmart to pay 100% of college tuition and books for associates. Press Release July 27, 2021. Retrieved from https://corporate.walmart.com/newsroom/2021/07/27/walmart-to-pay-100-of-college-tuition-and-books-for-associates on Sept. 12, 2021.

Wensing, M. (2021). Using supply chain predictive analytics to reduce risk. *Risk Pulse*, January 28, 2021. Retrieved from riskpulse.com, May 1, 2021.

Whiteside, Jeffrey. (2021). *Talent 2025*, presentation to Rotary Club of Evansville, March 2, 2021.

Wins, M. (2019). How are top fortune 500 companies using eLearning to boost productivity? *Supply Chain Academy Blog*, April 12. Retrieved from www.supplychain-academy.net, on sept. 2, 2021.

Wolf, Z. (2021) Why ending pandemic lockdowns created new shortages of ketchup and everything else. CNN Politics *What Matters*, Sun., April 11.

Business Analytics Categories

DECISION SCIENCES INSTITUTE
Examining the Mission Statements of the Most Ethical Companies

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

ABSTRACT

In this study, we performed text mining on the mission statements of the most ethical companies located in different countries across the globe. A SAS Enterprise Miner text analysis was conducted to explore if the most ethical companies in their mission statements have alluded to the same key terms and themes such as social responsibility and ethics. The purpose of this study, therefore, is to provide a systematic exploration of the mission statements of the most ethical companies in an attempt to identify themes and key terms contained within the mission statements

KEYWORDS: Text mining, most ethical companies, mission statement analysis.

INTRODUCTION

A mission statement is a general description of the purpose of the organization, its objectives, and why it exists. It often, in simple and concise terms, defines and communicates the core values, goals, and the primary purpose of an organization. While various factors may play an important role in defining and articulating a mission statement, a proper mission statement enriched with specific values is essential for effective communications (Law and Bresnik, 2018), and for formulating, implementing, and evaluating business strategy (Kemp and Dwyer, 2003).

Mission statements are examples of unstructured or textual data. Although they may vary in format, content, and length, one would expect to see some common terms and themes in mission statements across various organizations located in different countries, continents, and geographic regions. Therefore, textual trends, themes and patterns within mission statements may be extracted through text mining or text analytics and information and insights contained in the mission statements can be quantified.

This study aims at analyzing the mission statements of the most ethical companies across the globe in terms of how they describe their primary objectives, and define and communicate their core values, main purposes, and values. We conducted text mining using a commercially available text mining tool, SAS Enterprise Miner, to extract trends and patterns, quantify information contained in the mission statements, and survey if the most ethical companies shared any similarities with respect to some key terms such as social responsibility and ethics. The purpose of this study is to provide a systematic exploration of mission statements of the most ethical companies in an attempt to identify patterns of differences and similarities within these statements.

LITERATURE REVIEW

A mission statement may be defined as "'the operational, ethical, and financial guiding lights of companies," (Tian, 2004), which may be used as a strategic tool that emphasizes an organization's uniqueness and identity (Alegre et al, 2018). Kemp and Dwyer (2003) argue that a mission statement plays an important role in formulating, implementing, and evaluating business strategy.

A number of studies conducted in the past allude to the importance of mission statements. For instance, King et al. (2012) argue that mission statements are the most important organizational communication device to inform all stakeholders. Mas-Machuca et al., (2017) makes a similar observation claiming that mission statements are widely considered an effective strategic and communication tool in all types of organization. Moreover, Berbegal-Mirabent et al., (2021) argue that mission statements are a strategic tool that can provide a company with a purpose of being that communicates the core of the business to internal and external stakeholders.

Since they are used as an important communication tool, various studies in the past analyzed mission statements. For instance, Babnik (2014) analyzed the mission statements of 222 Slovenian companies from the organizational culture perspective. They found that mission statements play an important role in communicating organizational culture. Using a sample of 39 social enterprises located in Spain, (Berbegal-Mirabent et al., 2021) examined the link between mission statements and performance in social enterprises. They found that those firms for which the mission statement explicitly considers the customers and the product/service offer are more likely to exhibit higher economic performance. A study conducted by Mas-Machuca et al., (2017) examined the mission statements in a sample of 117 Spanish social enterprises and compared them with previous research on US-based social enterprises' mission statements. A similar study was conducted by Macedo et al., (2016) who surveyed 112 non-profit organization in Portugal. They conclude that the influence of organizational commitment should be considered to better understand the relationship between mission statements and organizational performance.

Some earlier mission statement analysis studies investigated the relationship between mission statements and firm performance. For instance, Baetz and Kenneth (1998) explored the possible relationship between mission statements and firm performance using a sample of 136 large Canadian organizations. Christopher (1997) took a similar approach to analyzing the content of industrial firm mission statements. The author concludes that some mission components matter more than others

The relationship between firm size, profitability, industry, and mission statements was a subject in a study conducted by Amato and Amato (2002). The authors state that financial services companies often include societal goals in their mission statements while metal, mining, and construction companies include safety and ecological goals. Kirk and Nolan (2010) took a look at the relationship between several attributes of mission statements in women's rights nonprofit organizations and organization financial performance. The authors claim that mission statements with a more focused geographic scope were associated with lower overhead ratios, whereas mission statements that focused on target client groups were associated with larger one-year increases in contribution. A study run by Williams (2008) investigated the mission statements of high performing and lower performing companies. The author concluded that the two groups of companies employed similar strategies for building corporate identities and images.

In a study conducted in 2011 by King et al., (2011), the mission statements of the 25 largest businesses in the U.S., Australia, Canada, and Great Britain were analyzed. The authors found that quality, core values, and leadership were the most frequently mentioned topics in the mission statements. In higher education, Palmer and Short (2008) analyzed the mission statements of 408 AACSB accredited schools with respect to the relationships between mission statement content and measures of business school characteristics, including performance. They found that business school mission lack comprehensiveness.

This study presents a systematic exploration of mission statements of the world’s most ethical companies located in different countries across the globe. In this study, we employed a text mining and analysis tool to extract trends and patterns and to quantify information contained in the mission statements.

RESEARCH METHOD

To analyze the mission statements of the most ethical companies, we used the data published by the Ethisphere Institute, the global leader in defining and advancing the standards of ethical business practices (worldsmostethicalcompanies.com). The Ethisphere Institute employs an assessment process that includes “more than 200 data points on culture, environmental and social practices, ethics and compliance activities, governance, diversity and initiatives to support a strong value chain. The process serves as an operating framework to capture and codify the leading practices of organizations across industries and around the globe.”

Having retrieved the list of the most ethical companies published by the Ethisphere Institute, we then visited each company’s web site and copied its mission statement into a text file. If no mission statement was available, other statements such as vision, goal, philosophy, a statement of purpose was used instead. Having compiled the mission statements into text file and uploaded it into SAS Enterprise Miner to run text analytics, the first node we used was the Text Parsing node, which is used to parse the text file. Next, we employed the Text Filter node to filter out the extraneous information and terms in the text file. Filtering helps to drop any terms that aren’t relevant for analysis and are nonessential for developing conclusions about the text file. The final node is the Text Cluster node, which is used to cluster the remaining terms from the text file after they have been parsed and filtered. The analysis that was run was a text analytics, which is also commonly known as text mining. The purpose of text mining is to attempt to extract meaningful information from text. This can be done systematically by a software application which automatically breaks down trends and patterns in order to draw insightful conclusions.

In the following section, we summarize and report our findings on the mission statements of the most ethical companies.

DATA ANALYSIS AND RESULTS

World’s Most Ethical Companies

The world’s most ethical companies examined in this study include 136 companies, span 22 countries, and employed a total of 7,675,162 employees in 2020. While the USA is home to the largest number of the most ethical companies, surprisingly, although it is the second largest economy in the world, no Chinese company made the list. Similarly, while Germany is home to some of the largest and most reputable companies in the world, no German company is included in this year's list.

Text Parsing and Analysis of the Mission Statements

We analyzed the mission statements using various Text Mining tools available in SAS Enterprise Miner to extract detailed information about the mission statements. During the analysis, the text mining nodes used were File Import, Text Parsing, Text Filtering, and Text Clustering.
We first used the Text Parsing node to parse text into smaller parts and to identify the most frequently appearing terms in the mission statements. The most important output for the Text Parsing Node is the Term Table (table 1). The individual records of the table describe various information about the terms that were collected during parsing process. The Frequency column in table 1 shows the total number of times any given term appears in the document containing the mission statements.

The term table (table 1) was sorted to show the top 10 terms based on their frequencies. As seen, such terms as "customer," "people," "health," and "world" are among the most frequently appearing terms in the mission statements of the world's most ethical companies.

Table 1: Most Frequently Occurring Terms

Term	Frequency
customer	33
people	32
health	29
world	27
life	23
energy	21
business	20
provide	19
technology	16
improve	14

Looking at table 1, it may be argued that like most companies, the most ethical companies focus on providing their "customers," with services and products they need, and "improving" "people's" "lives" and "business" through "technology"

Text Filtering and Identifying Significant Terms

While these terms are important to consider, greater value may be extracted from the mission statements by performing a further analysis known as "text filtering," which may allow one to see trends that may be present in the mission statements. Therefore, once we have identified the most frequently occurring terms in participating companies' mission statements, we then ran the Text Filter node, which is often employed to reduce and filter out extraneous information so that only the most relevant information can be extracted. The most relevant terms in the text corpus

or the importance of particular terms within a topic is identified by a weight computed by the term weighting method and assigned to each term in that text corpus.
Term weights are used the most important terms in a document collection. The assumption is that terms that are useful for categorizing documents are those that occur in only a few documents but they occur many times in those few documents, and those terms are significant in discriminating the documents (Chakraborty et al., 2013).

As tabulated in table 2, a completely different set of terms is identified as important across the mission statements of the most ethical companies. The term table (table 2) tabulates the top 10 most important terms based on their weights or importance and is sorted by the weight, largest to smallest. The terms at the top of this table have the greatest importance. Any term with a value for weight greater than zero is important, and all terms with weight equal = 0 have no importance. In other words. the value of weight determines the terms importance, a large weight represents a term has higher degree of importance and a lower weight represents a term is less important.

The weight of the terms is determined by TF-IDF; where
TF: Term Frequency=d (number of documents containing a given term)/D (the size of the collection of documents), and
IDF: Inverse Document Frequency IDF(t)=log (Total number of documents/Number of documents with term t in it).

This accounts for assigning important terms based on their importance to the document, and not just due to their frequency.

Top 5 most important terms are “social,” “ethics,” “responsible,” “nature,” and “member.” These terms have a weight of 0.762.

Table 2: Terms and Their Weights

Term	Weight
social	0.762
ethics	0.762
responsible	0.762
nature	0.762
member	0.762
planet	0.752
support	0.752
society	0.752
bank	0.752
tomorrow	0.752

A review of their mission statements yields interesting observations. For instance, some of the largest corporations such as 3M, Voya, and AVON speak of a “social responsibility” in their mission statements. Similarly, the term “ethics” holds significance for such well-known companies as Colgate-Palmolive. It appears, the most ethical companies strive to be “socially responsible," and ethical," support their "societies," and respect and protect "nature," "planet," and "environment."

Text Mining and Cluster Analysis

We performed a cluster analysis on the mission statements of the most ethical companies to create the smallest number of distinct clusters with highly similar records in the same clusters. The cluster analysis was run based on the Hierarchal algorithm. The end result was five clusters, each with a maximum of ten terms (table 3).

As seen, cluster 7 has the highest percentage with a frequency of 134 terms. It contains such important terms as "productive solutions," "person well-being," and "innovative," implying that the most ethical companies focus on providing productive and innovative solutions for the well-being of their customers. Similarly, clusters 6 and 10 speak of caring about the planet's health, how they are committed to serving their communities, improving lives through technology, and achieving sustainable planet and communities. Finally, cluster 4, which is the smallest cluster, speaks of the companies being committed to delivering quality services and products to their communities.

Table 3: Text Clusters

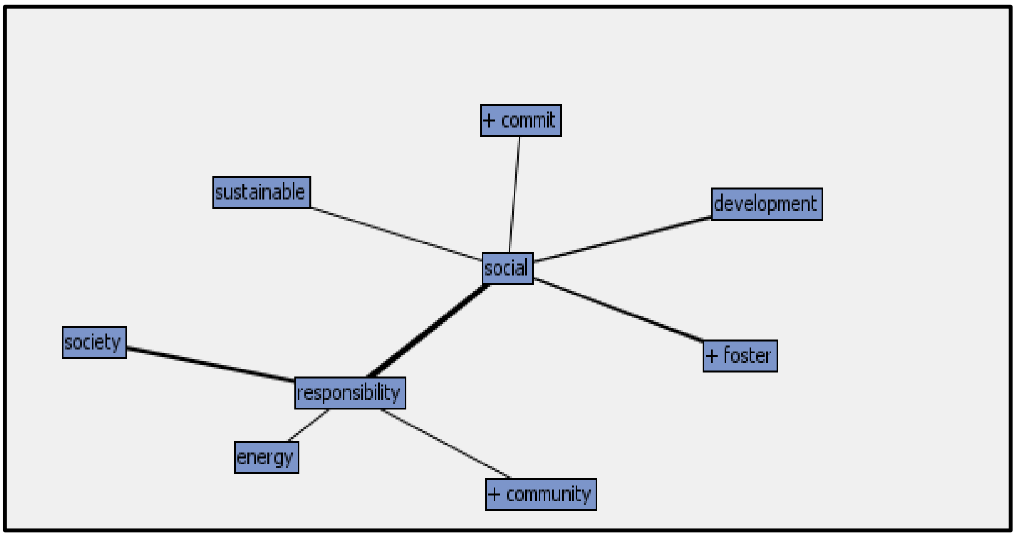
Cluster ID	Descriptive Terms	Freq.	%
4	+business services products +commitment +deliver quality communities +community +innovation financial	41	0.15
6	care +achieve planet health sustainable committed financial +company communities lives	48	0.18
7	brings founder productive solutions +person well-being +bring clients innovative +company	134	0.49
10	energy +technology +life +work group power services +improve +live +mission	48	0.18

Concept Links

We ran a further analysis to examine what terms are closely associated with one another in the mission statements of the most ethical companies. To explore the terms and their association, we constructed a "concept link diagrams" available in SAS Enterprise Miner. Concept link diagrams help in understanding the relationships between various terms based on the co-occurrences of the terms in the documents (Chakraborty, 2013). In other words, they are a visual representation of how the most important terms in a document are related to one another.

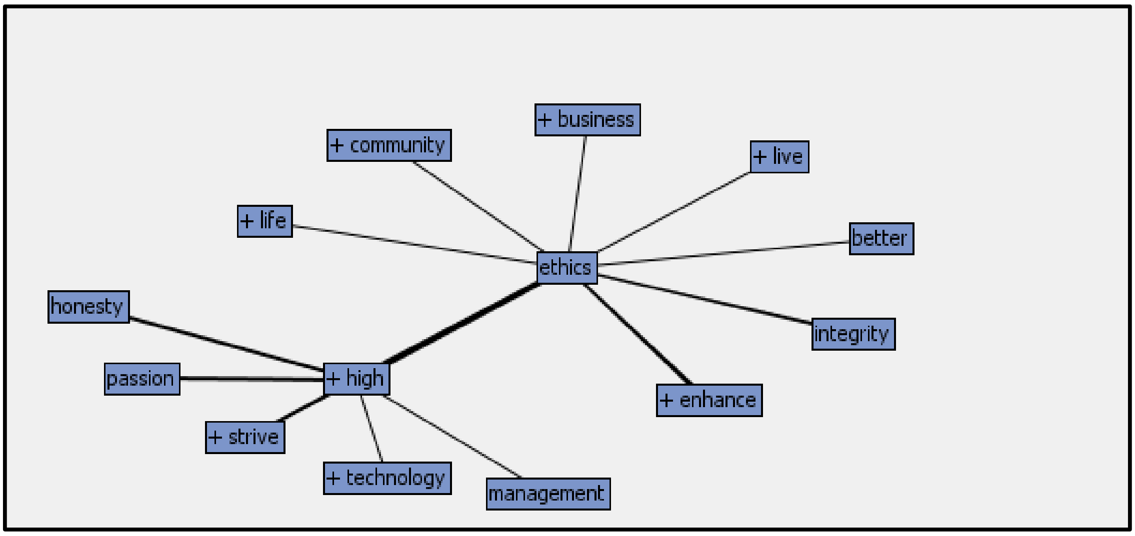
As seen in figure 1, the terms "sustainable," "development," "commit," "foster," and "responsibility" are significant terms that span off of the term "social." The terms "social" and "responsibility" are strongly and closely associated terms as indicated by the thickness of the line linking the two terms. We further expanded on the term "responsibility," As seen, it is strongly associated with the terms "society", "energy", and "community."

Figure 1: Concept linking for the term social



Another concept link diagram was created in SAS for the term “ethics,” which is the second most important term as summarized in table 2. Figure 2 shows the concept links for “ethics,” As seen, the terms “ethics” and “high” are strongly associated with one another. In addition, it is also associated with such important terms as “life,” “community,” “integrity,” “enhance,” and “live.” Similarly, the term “high” is closely associated with some important terms such as “honesty,” “passion,” and “strive**.”**

Figure 2: Concept links for ethics and high



The terms and the relationships among the terms in figure 2 yields a similar conclusion drawn in table 3 after running a cluster analysis. Both figures 2 and 3 suggest that the most ethical companies in their mission statements highlight the importance of being *socially responsible* and *ethical* while serving their *customers* and *communities* and *fostering* their *societies* through *sustainable business* practices. They further *strive* to *develop* their *communities*, *improve* and *enhance* their *customers*' *lives*, and practice *honesty* and *integrity* when serving their customers

DISCUSSION AND CONCLUSIONS

In this study, we examined the mission statements of the most ethical companies across the globe in terms of their main purposes, values, goals, and objective, and what they say about their vision and goals. We conducted text mining to investigate if the most ethical companies have valued the same key terms and themes.

It may be inferred that "social responsibility," and "ethics" are key areas of interest in the mission statements of the most ethical companies. Additionally, the terms "society," "nature," and "planet" hold statistical significance for these companies and their communication to their stakeholders and the public at large. A review of their mission statements indicated that some of the largest corporations surveyed in this study such as 3M, Voya, and AVON try to be "socially responsible," and ethical," support their "societies," and respect and protect "nature," "planet," and "environment." The world's most ethical companies that stress these weighted terms in their mission statements may do so to show their commitment by being socially responsible and ethical, and delivering sustainable business solutions to their customers.

A cluster analysis performed on the mission statements suggested that the most ethical companies focus on providing productive and innovative solutions for the well-being of their customers. Similarly, the cluster analysis revealed that such companies speak of caring about the planet's health, how they are committed to serving their communities, improving lives through technology, and achieving sustainable planet and communities. Moreover, their mission statements mention of being committed to delivering quality services and products to their communities.

The concept linking analysis, which identifies the most significant terms and the relationships among the terms, suggests that the most ethical companies in their mission statements highlight the importance of being socially responsible and ethical while serving their customers and communities and fostering their societies through sustainable business practices. They further strive to develop their communities, improve and enhance their customers' lives, and practice honesty and integrity when serving their customers.

This research is limited to the world's most ethical companies identified by the Ethisphere Institute. A follow up study may be conducted to compare the results of this analysis by breaking the data set into different geographic regions such as Europe, USA, and Asia. In addition, the same mission statements may be analyzed in terms of the industries to which such companies belong.

REFERENCES

Alegre I, Berbegal-Mirabent J, Guerrero A, Mas-Machuca M (2018) The real mission of the mission statement: a systematic review of the literature, *Journal of Management and Organization 24*(4), 456–473.

Amato, H. C, and Amato, H. L. (2002). Corporate commitment to quality of life: Evidence from company mission statements, *Journal of Marketing Theory and Practice*, *10*(4), 69-87.

Babnik, K., Breznik, K., Dermol, V. and Trunk Širca, N. (2014), The mission statement: organizational culture perspective, *Industrial Management & Data Systems, 114*(4), 612-627.

Baetz, B, and Kenneth, C. (1998). The relationships between mission statements and firm performance: An exploratory analysis, *Journal of Management Studies*, *35*(6), 823-853.

Berbegal-Mirabent, J., Mas-Machuca, M. & Guix, P. (2021). Impact of mission statement components on social enterprises' performance, *Review of Management Science*, 15, 705-724.

Chakraborty, G., Pagolu, M., and Satish, G. November (2013). *Text mining and analysis practical methods, examples and case studies using SAS*. SAS Institute. Cary, North Caroline.

Christopher, K.B. (1997). Industrial firms and the power of mission, *Industrial Marketing Management*, 26(4), 371-383.

Kemper, S and Dwyer L. (2003). Mission statements of international airlines: A content Analysis, *Tourism Management*, *24*(6), 635-653.

King, D., Case, C., and Premo, K. (2012). An international mission statement comparison: United States, France, Germany, Japan, and China, *Academy of Strategic Management Journal, 11*(2), 93-119.

King, D., Case, C., and Premo, K. (2011). A mission statement analysis comparing the United States and three other English-speaking countries, *Academy of Strategic Management Journal, 10*(51). 21-45.

Kirk, G., and Nolan, B. S. (2010). Nonprofit mission statement focus and financial performance, *Nonprofit Management and Leadership*, *20*(4), 473-490.

Law, K.M.Y., and Breznik, K. (2018) What do airline mission statements reveal about value and strategy?, *Journal of Air Transport Management*, 70, 36-44.

Mas-Machuca, M., Ballesteros-Sola, M., and Guerrero, A. (2017). Unveiling the mission statements in social enterprises: A comparative content analysis of us- vs. Spanish-based organizations, *Journal of Social Entrepreneurship, 8*(2), 186-200.

Macedo, M.I., Pinho, C., J., and Silva, M, A. (2016). Revisiting the link between mission statement and organizational performance in non-profit sector: The mediating effect of organizational commitment, *European Management Journal*, *34(*1), 36-46.

Palmer, B.T. and Short, C. J. (2008). Mission statements in US colleges of business: An empirical examination of their content with linkages to configurations and performance, *Academy of Management Learning and Education, 7*(4), 454-470.

SAS Institute Inc. 2015. Getting Started with SAS Text Miner 14.1. Cary, NC: SAS Institute Inc.

Tian, Y., 2004. *Being responsible to the society*. A *computer-assisted text analysis of the mission statements on the fortune 100 companies' web sites*, in proceedings of Annual Meeting of the International Communication Association, New York City, NY, May 20-25, 2004.

Williams, L, S. (2008). The mission statement: the corporate reporting tool with a past, present, and future, *International Journal of Business Communications*, *45*(2), 94-119.

DECISION SCIENCES INSTITUTE

Preventing and Mitigating Racial Bias in Machine Learning-Based Software

Bay Arinze
Drexel University
Email: arinzeob@drexel.edu

ABSTRACT

By 2023, roughly $100 billion will be spent on Artificial Intelligence (AI) systems, many making autonomous decisions in health, housing, employment, and other areas. Numerous AI systems incorporate bias from skewed learning samples and faulty training methods. Failing to remedy these problems will perpetuate and expand systemic discrimination, made more acceptable by being "objective computer decisions". The research aims to create safeguards against this practice. Past approaches to algorithmic bias include anecdotal storytelling, exhortations to AI professionals to "do better" and ad-hoc responses to complaints about specific systems such self-driving cars pedestrian recognition, AI-based redlining, disparities in the criminal justice system, facial recognition, and Facebook advertising, among many others. However, with a systemic problem, systemic solutions such as new methodologies, are necessary. The approach taken by this research is to target the methodologies for developing AI-based systems e.g., Cross-Industry Standard Process for Data Mining (CRISP-DM). Modifying the methodologies include steps to check how representative data samples are; instituting checks in model training methods; collecting additional data when minority populations are underrepresented; excluding race when it is irrelevant to the decision; and testing new AI systems for bias.

KEYWORDS: Machine Learning, Racial Bias, Software Development, Artificial Intelligence

INTRODUCTION

Machine learning algorithms and artificial intelligence-based software are increasingly used across many industries and organizational types in the United States. However, society is awakening to the inherent biases in many organizational applications created using these techniques, and by extension, the discriminatory actions that follow on from their use. In a country currently in the middle of widespread protests in support of racial justice, many of these computer systems arguably pose a greater danger than the more obvious ills and practices that these protests seek to curtail.

Propublica (2016) provided a compelling comparison of two crimes, one committed by an 18-year-old black woman and another by a 41-year old Caucasian man. In the first case, the woman stole a parked bicycle, while the male shoplifted $86 worth of tools. The white male had earlier been convicted of armed robbery and served 5 years in prison, and had another charge, while the black woman had several misdemeanors as a child. When being booked into jail, a computer program was used to predict the likelihood of each offender committing a future crime.

The white male was assigned a low risk, while the black female was assigned a high risk. This was not an academic exercise as these risk assessments are routinely used by judges to determine sentencing. Two years later, the black woman had not reoffended while the white male was serving 8 years for another robbery. A further systemic analysis of the software by the researchers showed inbuilt bias by comparing offenders' races, assessments, and subsequent rates of reoffending.

Bogen (2019) provides other examples of how candidate pools are shaped by bias in advertising positions. He writes: *"… targeted ads on Facebook for supermarket cashier positions were shown to an audience of 85% women, while jobs with taxi companies went to an audience that was approximately 75% black*." So even if efforts beyond initial screening are de-biased, which many are not, the initial damage is done by advertising selection. In addition, though employers have a legal duty to make their assessment tools de-biased, in practice, the subjective elements of candidate selection can very easily make their way into algorithms and software packages. This example and others are not isolated cases.

In a startling paper out of Georgia Tech, Wilson et al., (2019) found that people with darker skin are more likely than their white counterparts to be hit by self-driving cars, due to the nature of state-of-the art object detection models these cars use. The models were influenced by sample bias, i.e., not using enough minorities in the training process. Buolamwinin and Gebru (2018) similarly found that 3 commercial facial recognition packages had higher error rates for minorities, again, based on training the model using an unrepresentative sample. And of course, there is the infamous case back in 2015 of software engineer Jacky Alciné who pointed out that the image recognition algorithms in Google Photos were classifying his black friends as gorillas.

UNDERSTANDING BIAS IN AI-BASED SYSTEMS

Machine learning and other artificial intelligence techniques do offer great promise to organizations by emulating and often exceeding human skill in many activities previously thought to be the sole preserve of humans. However, there are now abundant examples to show how many of these systems in areas as diverse as the criminal justice system, hiring, facial recognition and AI-based automation may provoke another round of bias and discrimination against minorities, though much of it unintentional, through the use of biased algorithms.

There is growing awareness of the problem of biased algorithmic software, and specifically, software based on machine learning. This has resulted in some recommendations to rectify the damage that such systems can do to minorities and women. They include giving greater weights to underrepresented minorities in the samples such systems are trained on; increasing the samples of minorities in the data; testing such systems for racial biases after development; and checking to ensure that human-based and historical biases are not transferred into software that learns human expertise and historical data patterns. However, two areas are still lacking, namely a comprehensive methodology and framework for de-biasing algorithms and systemic policy recommendations for doing so.

This paper advances a comprehensive methodology for de-biasing machine learning-based applications that builds on current theory and best practices. Such a methodology must be easy to understand, practical, and demonstrably effective in helping developers apply de-bias their applications. The outlines of this methodology are outlined next.

POTENTIAL IMPACT OF BIAS

IDC forecasts worldwide spending on cognitive and artificial intelligence systems will reach $77.6 billion in 2022 and $97.9 billion in 2023, according to New IDC Spending Guide (IDC, 2019). This means that there will be tens of thousands of applications used in organizations that affect the lives of every person in the United States. If these often-autonomous systems exhibit even a percentage of the bias they are demonstrated to have, it will certainly lead to unfair, untenable, and indefensible conditions for racial minorities and another round of racial injustices.

The potential impact of this research effort is in helping to create a practical and effective methodology that can assist those developing software, especially autonomous software, that uses machine learning. This methodology could be applied in diverse fields, ranging from law enforcement, to medicine, to employment, to finance, to radically minimize the racial injustices that will undoubtedly occur as these systems proliferate.

Many organizations fail to understand that the issue is not only that that such biases are ethically and morally indefensible, but that they are also *bad* business. Biased systems lead to more expensive judicial mistakes, inferior candidates hired and/or promoted, and inordinate corporate exposure to liability, all of which are bad corporate (or governmental) outcomes that affect the bottom line. Thus, there is a dire need to not only disseminate knowledge of this phenomenon of biased software, but to provide solutions for developers to impact future software development as we become more and more a digital society.

APPROACH AND INNOVATION

The research approach is to modify the hugely popular 6-stage CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology for data analytics and machine learning development, specifically to prevent and mitigate racial bias. This will have the advantage of not having to create a separate framework for data analytics that has less chance of adoption, but instead adapt a framework that machine learning program developers are intimately familiar with.

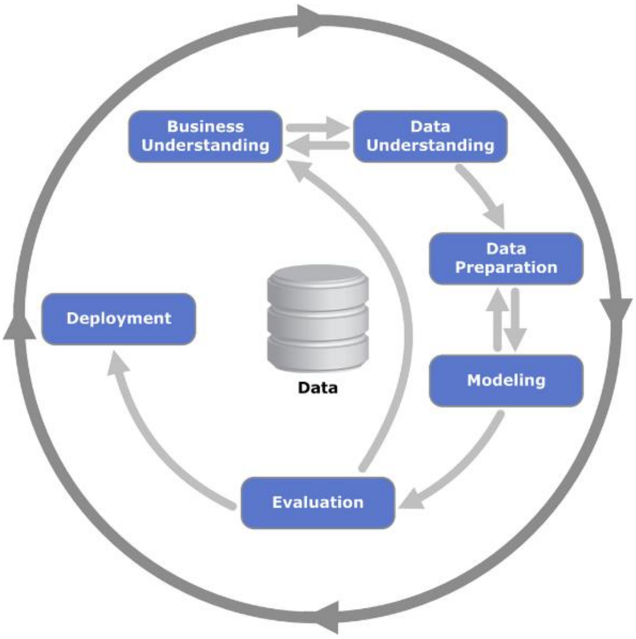


Figure 1: The CRISP-DM Data Mining Methodology

This is also the innovation of this approach, which is superior to scattershot guidelines that have been presented in multiple outlets as remedies to racial bias in software development. Instead, my approach will show *where* racial bias is to be evaluated in the development cycle, *where* it should be prevented mitigated and *how*, with enough use cases and detailed examples to be of practical use to developers.

Modifying CRISP-DM to Reduce Bias

Modifying CRISP-DM to reduce or eliminate bias will involve modifications to four of the six steps of the methodology. The result is a development approach for machine learning and other AI-based systems that is less biased intrinsically and more resistant to either conscious or unconscious injection of racial bias into the algorithms. The four recommended sets of modifications are as follows:

Data Understanding – Modifications

The data understanding stage involves the data scientists acquiring the data to be used for the project. The substages are:

a) Describe the data.
b) Verify data quality.
c) Generate a Data Quality Report.

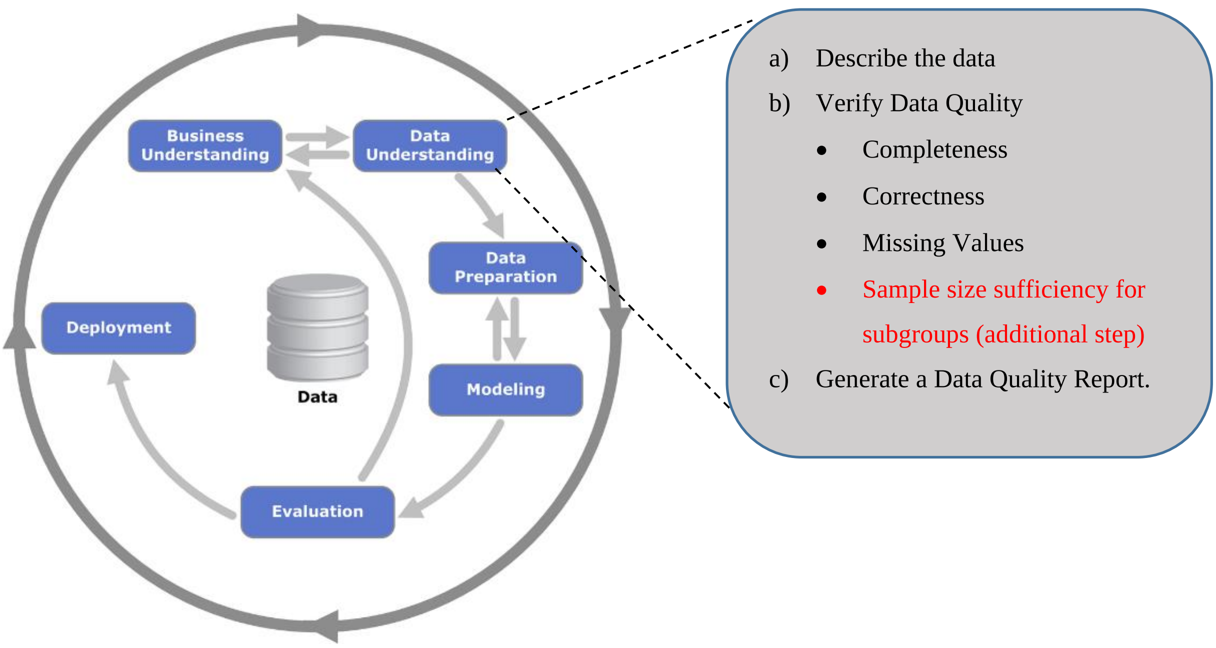


Figure 2: Modifying Stage 2 – Data Understanding (Additional steps in red)

In the *Data Understanding* stage, it is critical during verification of data quality to assess whether a sufficient sample size exists for all the difficult subpopulations that the system is intended to serve. This is where, if there are insufficient gender or racial sample sizes for training, an attempt is made to increase the sample size for these populations for training and testing purposes. A simple procedural reminder at this early stage in the machine learning process can itself serve as a powerful debiasing mechanism before the major efforts begin in a project.

Data Preparation – Modifications

In the *Data Preparation* stage (Figure 3), data scientists bring the acquired data to the level required by the modeling and analyses to be performed. The substages are:

a) Select the data needed.
b) Clean the data.
c) Construct e.g., transform parts of the data.
d) Integrate the data e.g., from multiple sources.

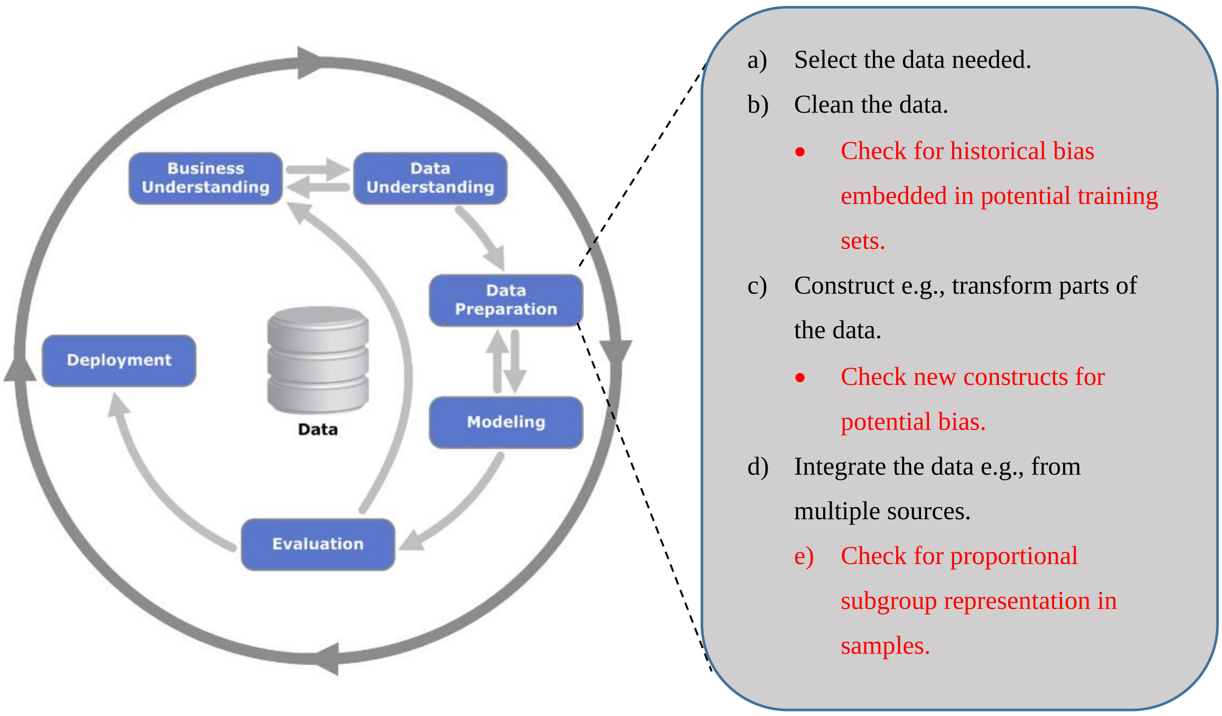


Figure 3: Modifying Stage 3 – Data Preparation

In the *Data Preparation* stage, as data is selected, cleaned and configured/reconfigured, it is also important to ensure that no implicit or explicit bias is introduced. To ensure this does not happen, three additional steps are proposed. As the data is cleaned, it should be inspected for potential historical bias as the outlines of training sets become clear. For example, as data is included or excluded, data scientists should ensure that subgroups are still proportionally represented- if not remedial action should be taken.

Second, when new constructs are being created from transformed data, developers should think about the potential bias of new constructs, particularly if they play a decisive role or have greater weights in future modeling. Examples could be zip codes for loans where historically, redlining has excluded specific groups. Finally, if data is integrated from multiple sources, developers should re-check sample sufficiency (from Stage 1) as these may have changed.

Modeling – Modifications
In the *Modeling* stage (Figure 4), the data scientists select modeling tools to be used for the machine learning/data mining project. The substages in modeling are:

a) Select modeling technique and assumptions e.g., normal/skewed distributions, missing values etc.
b) Generate a test design.
c) Build the model, including parameters.
d) Assess the model.

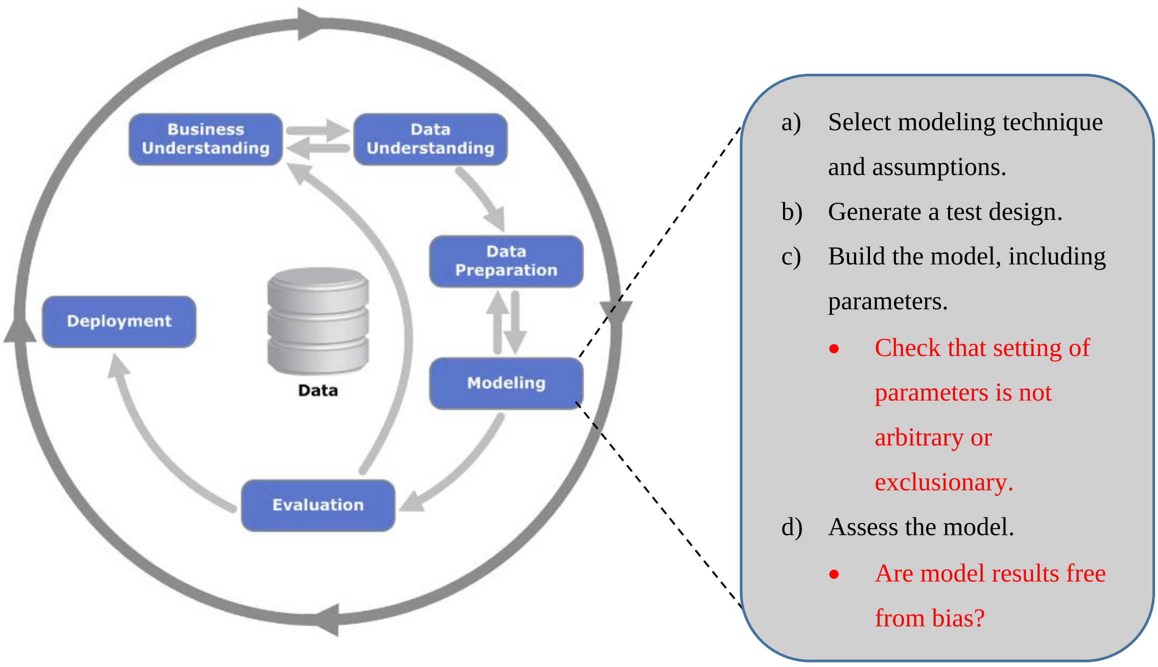


Figure 4: Modifying Stage 4 – Modeling

In the modified *Modeling* stage, there are also opportunities to ensure that the process is less prone to bias. Again, additional steps are proposed to achieve this. When the model is built, there are typically many parameters that must be introduced and/or adjusted. Do they put a finger on the scales?

For example, in some cases, perhaps race should be excluded because historical bias in training data meant that certain jobs were predominantly and unfairly giving to certain races, regardless of qualifications. Using race as a training variable would perpetuate the discrimination. Additional checks must also be instituted in the model assessment stage, going beyond typical measures like the confusion matrix, expected Value, AUC, ROC, and the like to determine whether models are free of bias.

Evaluation – Modifications
In the *Evaluation* stage (Figure 5), the data scientists review the overall process and determine if any factors or tasks have been overlooked. The substages in evaluation are:

a) Review activities and determine if there is a need for iteration.
b) Determine the next steps.

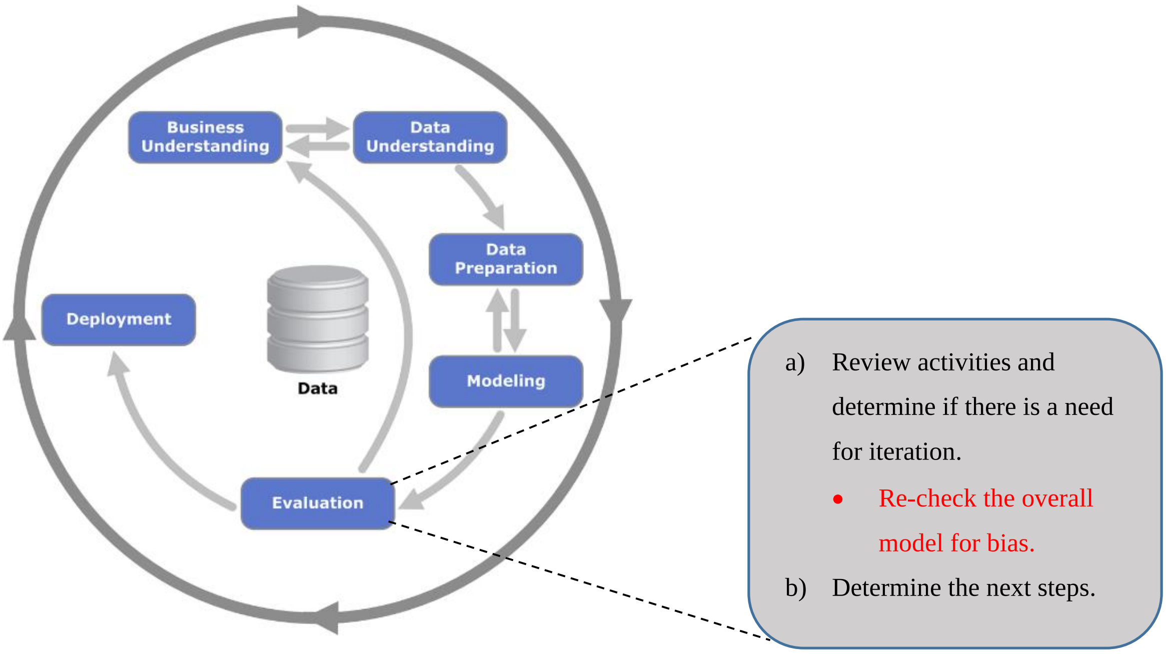


Figure 5: Modifying Stage 5– Evaluation

In the modified *Evaluation* stage, the model should undergo a holistic check again for bias. Using race as a training variable would perpetuate the discrimination. Additional checks must also be instituted in the model assessment stage, going beyond typical measures like the confusion matrix, Expected Value, AUC, ROC, and the like to determine whether models are free of bias.

DISCUSSION

Algorithmic racial biases are not just morally repugnant, they are bad business. Apart from institutionalized discrimination and unfairness in healthcare and housing etc., it leads to poorer business decisions, such as in hiring, firing, customer relationships, and even health outcomes. More AI-based decision-making can lead to a dystopian future, but this can be averted with improved methodologies used to create AI-based systems.

This paper has proposed a methodological addendum to CRISP-DM, the standard methodology for developing machine learning based systems. Used as suggested, it has strong potential to reduce or eliminate bias – conscious and unconscious – that is too-frequently introduced into machine learning models.

***Acknowledgements**: This work was funded by Drexel University's Racial Equity Rapid Response Research Awards Program*

REFERENCES

Angwin J, Larson J, Mattu S, Kirchner L., Machine bias: there's software used across the country to predict future criminals and it's biased against blacks, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Last visited 2017-12-17

Buolamwini J., and Gebru, T., Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, Proceedings of Machine Learning Research 81:1-15, 2018.

Wilson, B., Hoffman, J., and Morgenstern, J., Predictive Inequity in Object Detection, Georgia Institute of Technology, Feb 2019, arXiv:1902.11097.

IDC Corp, Worldwide Spending on Artificial Intelligence Systems Will Be Nearly $98 Billion in 2023, https://www.idc.com/getdoc.jsp?containerId=prUS45481219

DECISION SCIENCES INSTITUTE

Investigating Socioeconomic And Health Factors On The Prediction Of Life Expectancy By Using ML Techniques

Ozlem Cosgun
Information Systems Engineering and Management, Harrisburg University of Science and Technology, PA, USA, ocosgun@harrisburgu.edu

Megha Bhimrao Gaikwad
Information Systems Engineering and Management, Harrisburg University of Science and Technology, PA, USA, MBhimrao@my.harrisburgu.edu

ABSTRACT

In this paper, we investigated socioeconomic and health factors in the prediction of the life expectancy of the countries. Data for 193 countries were collected over 15 years between 2000 and 2015 from the WHO. Multiple Linear Regression, Decision Trees, Support Vector Machines, and Artificial Neural Networks (ANN) were utilized to predict and determine the significant variables on life expectancy. Then, we performed an information fusion-based sensitivity analysis to find the relative importance of the variables. Results show that adult mortality, income composition of resources, and schooling have higher effects on life expectancy in both developing and developed countries.

KEYWORDS: Life Expectancy, Machine Learning Techniques, Information-fused Sensitivity Analysis

INTRODUCTION

Life expectancy is the number of years a person may expect to live. It is determined by the year of their birth, their current age, and other demographic factors such as gender. It is a measure of a country's overall health and describes the average life span of a newborn. Problems such as war, illness, and poor health may reduce life expectancy. Life expectancy is the most important parameter for determining population health. Due to various advancements in public health, nutrition, and medicine, life expectancy has increased dramatically in the twentieth century. The most developed countries' life expectancy is expected to gradually increase before peaking in the mid-80s (Ediev, 2011).

Life expectancy is determined by a variety of factors, including health factors such as alcohol consumption, HIV, and a person's BMI. Even economic and social factors such as GDP, percentage expenditure, and education, among others, play a role. As a result, it is important to investigate certain causes to improve the nation's life expectancy. These factors have been examined in some previous studies (Mondal & Shitan, 2014; Meshram et al, 2021), however, in this study, we analyze 20 socioeconomic and health variables to predict which ones have the highest impact on the prediction of life expectancy and to identify the relative importance of the variables by using information fusion analysis. We also analyze what input variables show differences in developing and developed countries.

Some previous studies have considered only the effect of demographic factors, wages, and mortality on the life expectancy of the countries, but the health factors including the immunization and human development index were not taken into account which is very important in explaining the life expectancy (Mondal & Shitan, 2014). Some studies used several machine learning (ML) techniques and selected the best one in terms of their accuracy to decide on the significant variables (Meshram, 2021) that affect life expectancy. However, in this study, we first utilized several ML techniques to predict the life expectancy of developing and developed countries. Second, we compared the performances of the techniques in terms of RMSE and R2 values. Third, to find the relative importance of the variables in the prediction of the life expectancy, which is the main aim of this study, we utilized an information fusion-based sensitivity analysis method that combines the results of all machine learning models, instead of using the results of only one model. Then, the final analysis has been done to see if the variables having high impact on life expectancy are differ between developed and developing countries, which is the other aim of this study.

LITERATURE REVIEW

Life expectancy shows the general level of the population of any country. This gives the trend to infants, teenagers, and elderly individuals in all age groups. From 2000 to 2015, the average life expectancy has risen by 5.5 years (Roser et al, 2016). There are 80 years of life expectancy in the world's wealthiest countries. The life expectancy in Italy, Australia, Spain, and Switzerland was around 83 years in 2019 (Rojas-Rueda et al, 2019). However, life expectancy in Africa declined in the 1990s because of the AIDS epidemic. But it has now improved by 10.3 years to 61.2 years because of access to antiretroviral for the treatment of HIV (Nakagawa et al, 2013).

The share of the world's population over 60 years is projected to double from 11 percent to 22 percent, according to the WHO. A big effect on life expectancy is global aging and chronic disease (Mackenbach, 2002). For a while, life expectancy has been used as a measure of population health. The advantage of this is that it takes both the "quantity" and "quality" aspects of wellbeing in life and blends them. Life expectancy can be analyzed in various ways. In one of the studies (Mondal & Shitan, 2014), life expectancy has been measured in terms of inequality in the distribution of income. There is a high negative correlation between life expectancy and income inequality. Income inequality is not good for the nation and there are significant consequences that have been shown not for those with the lowest income but also higher-income individuals (Mackenbach, 2002).

The correlation between income disparity and mortality observed in aggregate studies could be due in part to improvements in academic performance (Lynch et al, 1998). The hypothesis with data for the US states has shown important connections between income inequality measures and age-adjusted mortality measures (Mackenbach, 2002). The main aim is to check whether various levels of formal education in the US have an impact on the relationship between income inequality and mortality. The authors used multiple regression analysis to analyze the census data from the years 1989 and 1990. Mortality is a dependent variable and Gini coefficient, per capita income, and percentage of people who are older than 18 years without a diploma are the independent variables (Mackenbach, 2002). Then, the results of regression have been improved by adding high school diploma to the regression model. So, education is a vital factor in the mortality of the nation which should be included in input variables.

Enhancing mortality and extending the quality of life requires timely and accurate data on levels and patterns of mortality. (Wang et al, 2016) includes cause-specific mortality for 249 reasons in 195 countries and territories from 1980 to 2015. The study (Wang et al, 2016) mainly includes an in-depth investigation of mortality patterns based on social demographics measures. In this article (Wang et al, 2016), they mainly used parameter selection for child and adult mortality synthesis by using Gaussian process regression to find out specific mortality because they are related to the Socio-demographics index. They also used various techniques and algorithms such as linear regression and socio-demographic index to analyze and predict the global, regional and national life expectancy based on mortality from 1980 to 2015. They also measured the magnitude of population growth and age structure. This shows that age-specific mortality has gradually increased over 35 years and this pattern has been continuously improved over the last decade. From the analysis, the life expectancy of some countries has decreased, and age-specific deaths have increased due to wars and diseases like HIV/AIDS in some countries (Nakagawa et al, 2013).

On the other hand, some authors (Nakagawa et al, 2013) analyzed the life expectancy of HIV-positive people using cohort studies and mathematical modeling studies, and due to the use of antiretroviral therapy, a healthy lifestyle, and testing of HIV, the life expectancy of HIV-positive people has been increased. So, the impact of population growth, aging, and improvements in age-standardized death rates globally varied significantly for various reasons. The authors conclude that, out of 25 variables, only 9 variables have an impact on life expectancy. They are HIV/AIDS, liver, total fertility rate, stroke, heart disease, and public health expenditure. Hence by improving those factors, the life expectancy of the regions or nations can be increased. Different from these studies, in this study, we used several machine learning algorithms such as SVM, ANN, ML, and DT to understand the effects of input variables on the target variable and to predict the life expectancy of the countries.

The health of the world's population has been improving gradually over the past 50 years. However, there are many health challenges all over the world. The analysis of health expenditure, gross domestic products (GDP), and education index has been done on the life expectancy of developing countries from 2006-2010 (Kabir, 2008). Cervellati & Sunde (2011) explored that the relationship between life expectancy and economic growth is non-monotonic. According to Kabir (2008), the country’s economic conditions are increasing due to the higher life expectancy rate. On the other side, according to authors (Cervellati & Sunde, 2011), the improvement in life expectancy causes increments in the population. Hence increment in population has a negative impact on income per capita. In addition, the effect of life expectancy on income per capita changes according to the different phases of economic and demographic development.

In addition, physical activities also reduce many diseases such as diabetes, heart diseases, stroke, cancer, etc. The mortality rate is decreased by about 30% to 35% because of physical activities (Ebrahim, 2011). According to the findings, those who engage in physical activities may expect to live 2 to 4 years longer (Mathers, 2004). Moreover, the study states that the life expectancy of aerobic endurance athletes is more than the of normal athletes (Mathers, 2004).

Some authors also discuss the effect of socio-economic status on life expectancy (Mathers, 2004). They conclude that the poor people living in developing countries face more problems, so they have lower life expectancy than the rich people in developed countries. The healthy life expectancy of African male is 40 years compared to female in developed countries who has 70 years of health expectancy in the year 2000. So, the life expectancy of a person with a low socio-economic status is lower. Guralnik et al (1993) have studied the socioeconomic status of disability-free or active life expectancy of black and white people of age 65 and more in Carolina, USA. They used disability as an independent variable such as walking, bathing. Sex, education, and race have also been taken for subgroups. The data were used to calculate the life expectancy, healthy life expectancy, and disabled life expectancy.

Based on the existing literature, there are some gaps observed and we provide the following contributions to the existing body of knowledge in this area:

- First, the effect of income inequality has been studied only in the United States (Lynch et al, 1998). They support the belief that there is a gradual depletion of evidence for a link between income inequality and population health and this has been seen only in the United States. However, in our study, we analyzed the effects of income inequality for all developed and developing countries.
- Second, most of the studies are generally based on 3 factors including mortality, education, and income (Rogot,1992). Nakagawa et al. (2013) only discuss healthy life expectancy in a few developing and developed countries and considers only HIV and life expectancy at birth. But, in our study, we consider 20 factors including mortality, schooling, income composition of resources, GDP, BMI, some diseases, total expenditure, etc. that affect the life expectancy of the nation.
- Third, the previous studies (Cervellati & Sunde, 2011; Acemoglu & Johnson, 2007; Ebrahim, 2011; Mathers, 2011; Kabir, 2008) only used a theoretical identification strategy to analyze the data and the data is not evaluated using any experiments or techniques. Meshram (2021) used ML techniques to predict the life expectancy of the countries but selected the one which has the highest accuracy to determine the significant variables. However, we used information-fusion bases sensitivity analysis that combine the results of several ML techniques and found the relative importance of the variables on the life expectancy.
- Finally, we made a comparative analysis to see what variables show differences in the life expectancy of the developing and developed countries.

METHODOLOGY

In this section, the research methodology of the study is described. First, we present the stages of data collection and then give information about the ML methods used to predict the life expectancy of the countries. After we present performance metrics used for the comparison of machine learning techniques, information-fusion sensitivity analysis is described.
We use Excel and R for the preparation of data and the analysis of data for predictive modeling.

Data Collection and Preparation

Secondary data is used in this study. The dataset is acquired from World Health Organization. It includes the data from the year 2000 to 2015. Data preprocessing is so crucial for the accuracy of the models, so the data is cleaned from missing variables and eliminated from the unrelated predictors and multiple entries. So, finally, we have about 2093 observations with 22 variables including socioeconomic and health factors, related to 195 countries. The data is shown in Table 1.

Table 1: Variables/factors considered in the study.

Factors/Variables	Description	Data Types
Country	Country observed	Nominal
Year	Year Observed	Numeric
Status	Developed or developing country	Nominal
Life Expectancy	Life Expectancy in age	Numeric
Adult Mortality	Probability of dying between 15 and 60	Numeric

	per 1000 population	
Infant deaths	No. of infant deaths per 1000 population	Numeric
Alcohol	Alcohol recorded per capita	Numeric
Percentage. expenditure	Expenditure on health as a percentage of GDP	Numeric
Hepatitis B	Hepatitis immunization coverage	Numeric
Measles	No. of reported measles	Numeric
BMI	Average Body mass index of entire population	Numeric
Under. Five. Deaths	No. of under-five deaths	Numeric
Polio	Polio immunization coverage	Numeric
Total Expenditure	General government expenditure on health	Numeric
Diphtheria	Diphtheria tetanus immunization coverage	Numeric
HIV	Deaths per 1000 live births	Numeric
GDP	Gross Domestic Products per capita	Numeric
Population	Population of the country	Numeric
Thinness. 5-9 years	Prevalence of thinness among children for age 5 to 9 %	Numeric
Thinness. 10-19 years	Prevalence of thinness among children and adults for age 10 to 19 %	Numeric
Income composition of resources	Human Development Index in terms of income composition of resources	Numeric
Schooling	Number of years of schooling	Numeric

Machine Learning Techniques

We use several Machine Learning techniques to find what variables have a high impact on life expectancy and predict the life expectancy of the countries. Our target variable is life expectancy which is a continuous variable. So, we employ regression-based techniques consisting of Artificial Neural Networks, Support Vector Machines, Decision Trees, and Multiple Regression in this study. In this section, we briefly give information about these techniques.

<u>Artificial Neural Networks</u>

Neural networks are made up of neurons, which are simple input/output components. These input/output units are connected, and each relationship has a weight assigned to it. The versatility of Neural Networks allows them to be used for regression and classification. Since our target variable is continuous, we used a Neural Network for regression in this analysis. Regression is used to define a relationship between a dependent variable and one or more independent variables. Although Neural Networks are complex and computationally expensive, they are adaptable and capable of selecting the best regression method (Hopfield, 1988). To feed the data into Neural Networks, we used the min-max approach to normalize the dataset.

<u>Support Vector Machines</u>

Support Vector Machines (SVMs) are supervised learning models for data classification and regression analysis. The SVMs employ Support Vector Regression (SVR) to estimate the continuous variable. The epsilon-insensitive tunnel is where SVR matches the best line within a set of values. By specifying an acceptable error margin (epsilon-insensitive tube) and the tolerance for falling beyond the error margin, SVR essentially allows us to choose how sensitive our model is to errors (Gunn, 1998).

Decision Trees

Data is continuously divided based on a parameter in Decision Trees which is a form of Supervised Machine Learning techniques. The tree is depicted using decision nodes and leaves. The data is divided at the decision nodes, and the leaves reflect the decisions or outcomes (Karacan et al, 2020). There are some popular decision tree algorithms including C4.5, C5, Classification and Regression Trees, etc. C5 algorithm was used in our study.

Multiple Linear Regressions

Multiple Linear Regression (MLR) is the most common form of linear regression analysis. It is a statistical analysis that describes the relationship between the dependent variable and two or more independent variables by fitting a linear equation to the observed explanatory variables. The best line is calculated by using the least-squares model that minimizes the sum of squares of deviations from the actual data (Tranmer & Elliot, 2008). We used MLR to estimate the life expectancy of the countries based on all predictor variables in this analysis.

k-fold Cross-Validation

We performed 10-fold cross-validation to check the accuracy of the models. Cross-validation is a resampling technique for evaluating machine learning models on a small sample of data. The whole data is divided into *k* equal folds in which one is used for testing while the others are used for training, The process is run *k* times and each *k* fold is used only once for validation. Then the overall accuracy is calculated by taking the average of *k* results to make a single estimation (Rodriguez & Lozano, 2009), (eq. 1):

$$Cross\ Validation\ Accuracy\ (CVA) = \frac{1}{k}\sum_{i=1}^{k} Accuracy_i \tag{1}$$

where *k* denotes the number of folds which is 10 in our case.

Performance Metrics and Model Evaluation

In this study, 80% of the data is used for the training data to find the relationship between output and input variables, and 20% is used for testing to assess the performance of the model. The accuracy and the performance of the model were compared according to the coefficient of determination (R2) value with the root mean square error value (RMSE) of each predictive model. R2 is a measure of the proportion of variance in the outcome which can be explained by predictor variables. The R2 is also the correlation between observed output values and the model’s expected values. So, the model is stronger when the R-squared is higher. In addition, the Root Mean Squared Error (RMSE) is a metric that measures the model's average error in predicting an observation's outcome. So, high R2 with low RMSE is acceptable and works well to predict the output variable.

Information fusion-based Sensitivity Analysis

Information fusion combines the results obtained from more than two models (Delen et al, 2017). A prediction model can be denoted as:

$$\hat{y} = g(x_1, x_2, \dots, x_n)$$

where $\hat{y}$ denotes the output variable and $x_1, x_2, \dots, x_n$ denotes the input variables. A fusion model which combines the m prediction models is written as:

$$\hat{y}_{fused} = \gamma\left(\hat{y}_{individual,i}\right) = \gamma\left(g_1(x_i), g_2(x_i), \dots, g_m(x_i)\right)$$

where γ is the operator that combines all individual prediction models (Oztekin et al, 2013).

Then we used sensitivity analysis to rank the importance of the variables in prediction. It is an experimental process that analyzes the impact of each variable on the output variable by removing it from the model (Oztekin et al, 2013). This is represented by the following formula:

$$S_i = \frac{V(i)}{V(F_t)} = \frac{V(E(F_t|x_i))}{V(F_t)}$$

where S_i denotes the importance of variable i, $V(F_t)$ denotes the variation in the output variable and the nominator refers to the variance of the expected value of all variables except x_i.

As expected, all machine learning techniques provided different results and the variables have different significances. So, to combine these results and produce one single fused result, sensitivity analysis was used. Sensitivity measure of variable *n* with information fused by *m* prediction models is given by:

$$S_{n(fused)} = \sum_{i=1}^{m} w_i S_{in} = w_1 S_{1n} + w_2 S_{2n} + \cdots + w_m S_{mn}$$

where w shows the standardized weight values of each machine learning model and S_{in} denotes the sensitivity measure of the n^{th} variable in the i^{th} model ($i \in m = 4 \text{ machine learning models in our case}$).

RESULTS

The performance of the algorithms for developed and developing countries are obtained and given in Table 2 and 3 respectively. Tables show the root mean square error (RMSE) and R2 for each machine learning technique used in this study.

Table 2: Performance of the Machine Learning Techniques for developed countries

Machine Learning Technique	RMSE	R2
Multiple Regression	0.068	0.830
Decision Tree	0.100	0.700
Artificial Neural Network	**0.055**	**0.921**
Support Vector Machines	0.095	0.782

Table 3: Performance of the Machine Learning Techniques for developing countries

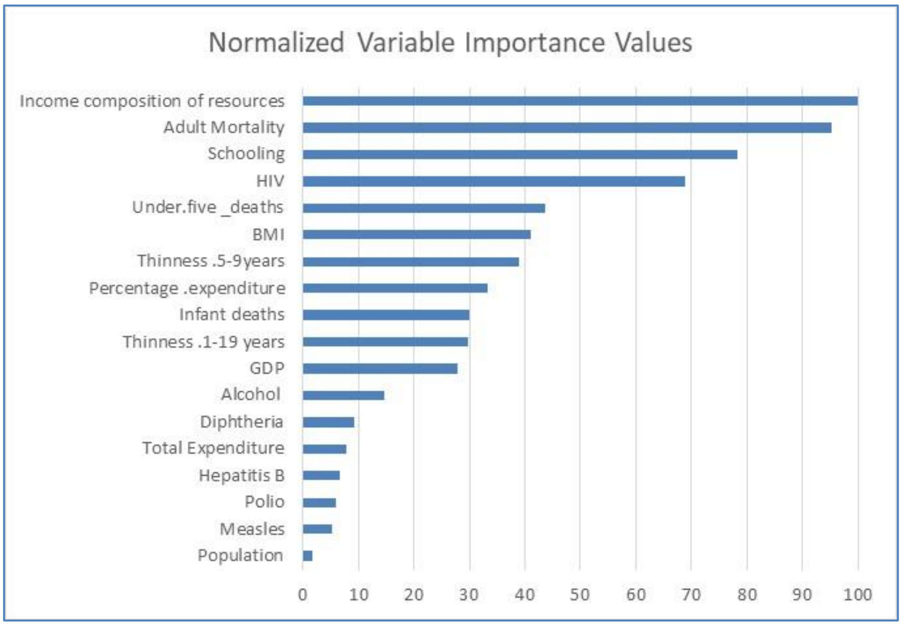
Machine Learning Technique	RMSE	R2
Multiple Regression	0.072	0.810
Decision Tree	0.200	0.620
Artificial Neural Network	**0.064**	**0.834**
Support Vector Machines	0.084	0.715

As seen from Tables 2 and 3, ANN shows the best performance to predict the life expectancy of the countries in terms of both RMSE and R2. The next best one, which has the highest performance, is Multiple Regression, and then SVM and DT follow these techniques, respectively.

However, our main intention is to identify the significant variables that have the highest impact on life expectancy and their rankings for developed and developing countries. So, we performed information fused sensitivity analysis for each country group to merge the results and decide on the single fused significant variables on life expectancy. The normalized variable importance values are given in Figure 1 and Figure 2.

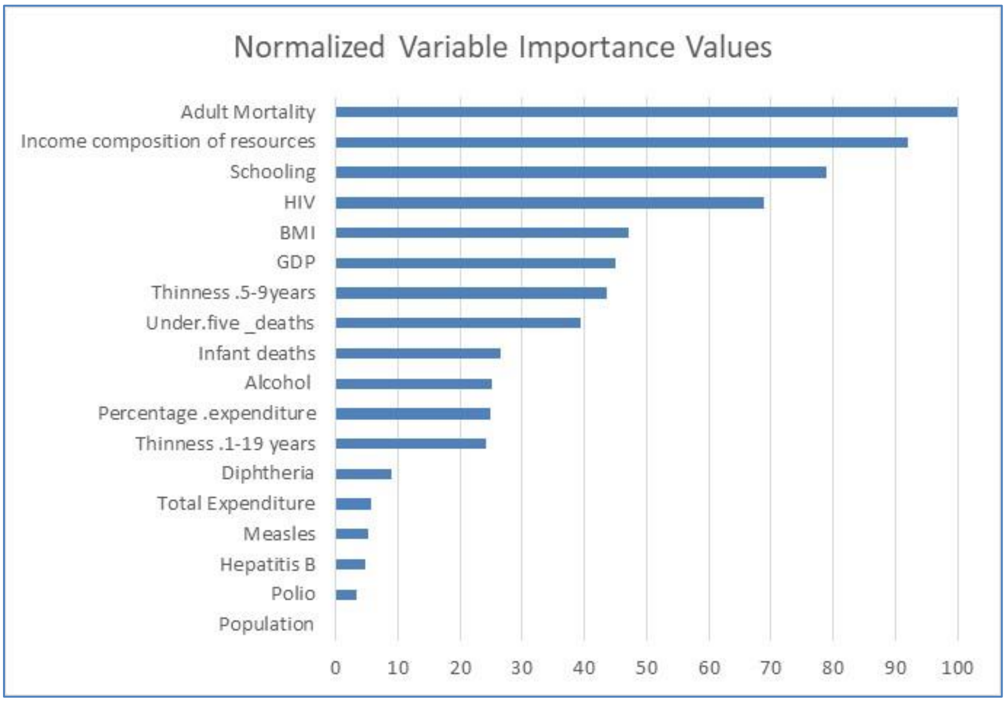
Sensitivity analysis results show that adult mortality, income composition of resources, schooling, and HIV have the highest impact on the life expectancy of both developed and developing countries. Under-five deaths, BMI, thinness 5-9 years and percentage expenditure follows these top variables in developed countries, and BMI, GDP, thinness 5-9 years, and Under-five deaths follow in developing countries. However, population and diseases including Polio, Measles, Hepatitis B have the least significance in the prediction of life expectancy of the countries.

Figure 1: Normalized Variable importance values for developed countries



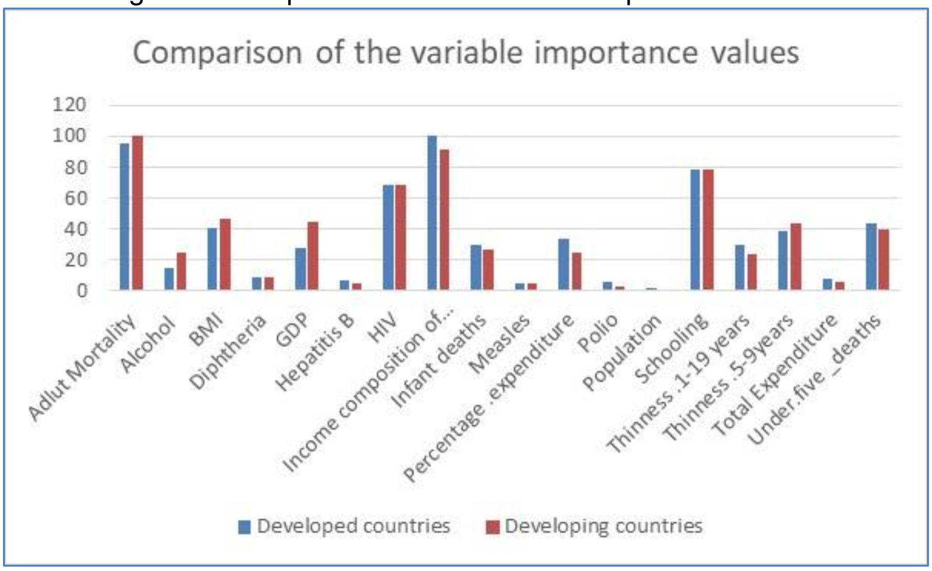
Cosgun, Gaikwad Investigating Socioeconomic and Health Factors

Figure 2: Normalized Variable importance values for developing countries



The comparison of the variables' importance values for developed and developing countries is also given in Figure 3. The variables have almost the same importance values except for GDP, Alcohol, and percentage expenditure variables. GDP and alcohol have a higher impact on the life expectancy of developing countries than developed countries, and the percentage of expenditure has a higher effect in developed countries than developing countries.

Figure 3: Comparison of the variables' importance values



DISCUSSION AND CONCLUSIONS

This study aimed to make the most accurate predictions about life expectancy and determine which variables have a higher impact on predicting the life expectancy of the countries. In this research, we looked at how socioeconomic and health factors could be used to predict life expectancy in both developing and developed countries. We used several machine learning techniques including ANN, SVM, DT, and MLR to predict life expectancy and determine the impact of variables on the life expectancy of the countries. Then, we performed an information-fused sensitivity analysis to combine the results of all these machine learning techniques to rank the importance of the variables in prediction. Based on the sensitivity analysis, adult mortality, income composition of resources, schooling, and HIV have the highest impact on the life expectancy of both developed and developing countries and most of the variables have similar variable importance values for these countries except GDP, alcohol, and percentage expenditure variables. Results show that our developed model provides significant insights for the viability of machine learning techniques in the determination of variables influencing the life expectancy of the countries.

REFERENCES

Acemoglu, D., & Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of political Economy*, 115(6), 925-985.

Cervellati, M., & Sunde, U. (2011). Life expectancy and economic growth: the role of the demographic transition. *Journal of economic growth*, 16(2), 99-133.

Delen, D., Tomak, L., Topuz, K., Eryarsoy, E. (2017). Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. Journal of Transport & Health, 4, 118-131.

Ebrahim, S., Taylor, F., Ward, K., Beswick, A., Burke, M., & Smith, G. D. (2011). Multiple risk factor interventions for primary prevention of coronary heart disease. *Cochrane database of systematic reviews*, (1): CD001561.

Ediev, D. M. (2011). Life expectancy in developed countries is higher than conventional estimated. Implications from improved measurement of human longevity. *Journal of Population Ageing*, 4(1-2), 5.

Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits and Devices Magazine*, *4*(5), 3-10.

Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS technical report*, *14*(1), 5-16.

Guralnik, J. M., Land, K. C., Blazer, D., Fillenbaum, G. G., & Branch, L. G. (1993). Educational status and active life expectancy among older blacks and whites. *New England Journal of Medicine*, 329(2), 110-116.

Kabir, M. (2008). Determinants of life expectancy in developing countries. *The journal of Developing areas*, 185-204.

Karacan, I., Sennaroglu, B., & Vayvay, O. (2020). Analysis of life expectancy across countries using a decision tree. *Eastern Mediterranean Health Journal*, 26(2), 143-151.

Lynch, J. W., Kaplan, G. A., Pamuk, E. R., Cohen, R. D., Heck, K. E., Balfour, J. L., & Yen, I. H. (1998). Income inequality and mortality in metropolitan areas of the United States. *American journal of public health*, 88(7), 1074-1080.

Mackenbach, J. P. (2002). Income inequality and population health: evidence favoring a negative correlation between income inequality and life expectancy has disappeared. *BMJ*, 2002; 324.

Mathers, C. D., Iburg, K. M., Salomon, J. A., Tandon, A., Chatterji, S., Ustün, B., & Murray, C. J. (2004). Global patterns of healthy life expectancy in the year 2002. *BMC public health*, 4(1), 1-12.

Meshram, S.S. (2020). Comparative Analysis of Life Expectancy between Developed and Developing Countries using Machine Learning. IEEE Bombay Section Signature Conference.

Mondal, M. N. I., & Shitan, M. (2014). Relative importance of demographic, socioeconomic, and health factors on life expectancy in low-and lower-middle-income countries. *Journal of Epidemiology*, 24(2), 117-124.

Nakagawa, F., May, M., & Phillips, A. (2013). Life expectancy living with HIV: recent estimates and future implications. *Current opinion in infectious diseases*, 26(1), 17-25

Oztekin, A., Delen, D., Turkyilmaz, A. & Zaim, S. (2013). A machine learning-based usability evaluation method for eLearning systems. *Decision Support Systems*, 56, 63-73.

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2009). Sensitivity analysis of k-fold cross-validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence*, 32(3), 569-575.

Rogot, E., Sorlie, P. D., & Johnson, N. J. (1992). Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. *Public health reports*, 107(4), 457.

Roser, M., Ortiz-Ospina, E. & Ritchie, H. (2016) - "Life Expectancy". Published online at OurWorldInData.org.

Tranmer, M., & Elliot, M. (2008). Multiple linear regression. *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, *5*(5), 1-5.

Wang, H., Naghavi, M., Allen, C., Barber, R. M., Bhutta, Z. A., Carter, A., & Bell, M. L. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The lancet*, 388(10053), 1459-1544.

DECISION SCIENCES INSTITUTE

Multi-Cloud Computing Adoption: A Conceptual Model

Md Golam Kibria
Department of Information, Operations and Technology Management
John B. and Lillian E. Neff College of Business and Innovation
The University of Toledo, Toledo, Ohio, USA
Md.Golam.Kibria@utoledo.edu

Marcelo J. Alvarado Vargas
Department of Management
John B. and Lillian E. Neff College of Business and Innovation
The University of Toledo, Toledo, Ohio, USA
Marcelo.AlvaradoVargas@utoledo.edu

ABSTRACT

Multi-cloud computing is an emerging technology has drawn increasing attention from both academics and practitioners. The adoption of multi-cloud computing brings tremendous business benefits for organizations. However, the deployment of multi-cloud computing is very challenging because of its newness, relational nature between parties and security concerns and therefore, a critical analysis is needed to examine the relevant aspects that influence the adoption of multi-cloud computing. This paper presents a conceptual model for multi-cloud computing adoption from different theoretical lenses. The paper also presents implications for practitioners and future research directions for researchers.

KEYWORDS: Multi-cloud adoption, management fads and fashions, technology mindfulness, trust, Relational IT capabilities, and Organizational IT capabilities.

INTRODUCTION

Multi-cloud systems use multiple cloud sources to provide services on-demand sequentially or simultaneously (Frank & Jones, 2019).The multi-cloud computing has been widely accepted as a new paradigm to manage information technology (IT) resources. It is a source innovation that introduces new business models and services to organizations (He & Wang, 2015). Moreover, the growth of emerging technologies including the internet of things (IoT), artificial intelligence and blockchain has pushed the adoption of multi-cloud a step ahead. Consequently, perceiving the value of multi-cloud computing, firms are interested in investing in this new technology (Frank & Jones, 2019). In a study, Frank and Jones (2019) found that more than 90 percent of IT leaders believe that majority of their services will be delivered through multi-cloud environment by 2021. This transition to multi-cloud environment from traditional in-premise data center is a generational transition for IT and therefore, it requires a close attention to know how to adopt and implement the multi-cloud environment successfully. The rationale behind using multi-cloud computing by firms is to leverage the benefits of the different cloud providers and to maximize their utility in single network architecture.

The multi-cloud adoption is an integral part of digital transformation for firms as it provides the opportunity to get on-demand services with flexibility and limitless scale. Although many organizations realized the value of multi-cloud computing adoption, they do not know how to adopt and implement it in organizational settings (Frank & Jones, 2019). However, the adoption of multi-cloud is not an easy task as it is a new technology as well as it involves multiple organizations. It will be undoubtedly a risky move if not adopted properly. Therefore, it is important to explore the key factors that facilitate or inhibit the adoption of multi-cloud computing. The previous studies in multi-cloud computing are mostly descriptive in nature (He & Wang, 2015; Lahmar & Mezni, 2018; Petcu, 2013). For example, Ardagna (2015) stated the current challenges and future applications of multi-cloud computing, whereas Hong et al. (2019) presented the current status of multi-cloud computing. No prior study is found that explores the determinants of multi-cloud computing adoption. Therefore, our study will be the first, to best of our knowledge that will explore the critical factors of adopting multi-cloud computing.

As we can assume a substantial growth of multi-cloud computing use in future, IT leaders should critically examine how they adopt multi-cloud computing in organizational setting. Failing to consider the antecedents of multi-cloud adoption, firms may risk themselves into technologies they don’t want to use and exposing themselves to unnecessary risk. We firmly believe that by identifying and analyzing the factors, the factors with the greatest effect on the adoption can be isolated. As a result, administrators should concentrate on these essential variables and gradually refine them in order to greatly boost the performance of multi-cloud adoption.

LITERATURE REVIEW

Extensive research has been conducted on information technology adoption in the field of information systems (IS) at both individual and organizational levels in various contexts (e.g. developing countries, education, and health care areas). Hence, research on information technology adoption is now the most mature stream of IS research. However, the existing technology adoption theories are old fashioned and have limitations in explaining the adoption of emerging information technologies (Chang et al., 2018; Sun & Jeyaraj, 2013). Therefore, the new innovative technology should be studied from beyond the dominant paradigm perspective (Fichman, 2004).

Since the multi-cloud is a new phenomenon in IS research, the most studies are descriptive, exploratory and case study based research (Ardagna, 2015; Gundu et al., 2020; He & Wang, 2015; Hong et al., 2019; Petcu, 2013). For example, Khajeh-Hosseini et al. (2012) developed a conceptual model for supporting cloud computing adoption decision. They qualitatively found the importance of costs, customer relationships, public image and flexibility on cloud adoption. Many more studies descriptively explained the multi-cloud adoption issues (Ardagna, 2015; Hong et al., 2019; Khajeh-Hosseini et al., 2012; Lahmar & Mezni, 2018; Petcu, 2013). Moreover, most published papers focused on technical aspects of multi-cloud adoptions. For example, He and Wang (2015) proposed a hybrid cloud model that integrates private and public clouds in a single architecture. They also illustrated the utility of the model based on two case studies. We found another prolific aspect of multi-cloud computing in IS literature that a substantial number of articles paid attention to trust issues in the adoption stage (AlZain et al., 2012; Bohli et al., 2013; Sengupta et al., 2011). This trust issue is brought by the lack of security in cloud computing implementation (Sengupta et al., 2011).

Our extant literature search revealed that there is dearth of research in multi-cloud adoption. The most studies focused on the overview, current challenges and opportunities, technical and security issues. Moreover, we found that most of the previous cloud adoption studies do not have grounded theories to guide their research, and therefore, this paper aims to explore the determinants of multi-cloud adoption based on the relevant theories.

THEORIES AND PROPOSITIONS DEVELOPMENT

Multi-cloud computing is a relatively recent topic in the area of information systems. Being a new technology, we can consider multi-cloud computing as a new innovation (Rogers, 1983). The adoption of an innovation largely depends on firm's management. Thong (1999) quantitatively found that firm's management such as CEO has significant influence on any technology innovation. Hamel & Prahalad (1994) argue that the firm with high technology orientation has superior performance over its competitors in rapid changing technological environment. In sum, we can conclude that the innovation adoption decision is made by the management of an organization. Therefore, the decision to adopt multi-cloud computing is contingent on management interests. Hence, we believe that the theory of management fads and fashions can explain the behavioral intention (BI) of multi-cloud adoption.

The theory of management fashion in business and management studies states that, under conditions of uncertainty, organizations ("management fashion followers") imitate innovation models promoted by "fashion-setting organizations" (such as: consulting firms; management gurus; business mass-media publications; and business schools) and that the diffusion rates and final levels of adoption of any given management innovation cannot be fully explained by rational/technically efficient arguments (the "efficient choice" perspective) (Abrahamson, 1991). In addition to techno-economic forces, socio-psychological factors have a significant influence in decisions to adopt and engage in continued use of a management innovation. This management fads and fashions has two dimensions: fashion broadcast and faddish contagion (Piazza & Abrahamson, 2020). The fashion broadcast refers to the broadcasting the theorizations of management innovations by professionals, entrepreneurs and fashion settlers in the industry. In the multi-cloud computing context, we argue that when a firm adopts the cloud technology, it sends a signal to the industry that an innovative technology has been implemented. Consequently, the management of other firms realize potentiality of the technology, which may develop a desire to adopt the technology. Faddish contagion, on the other hand, refers to conceptualization of the broadcasted practices and causes the adoption of these technologies (Piazza & Abrahamson, 2020). In multi-cloud context, we believe that a firm intends to adopt a technology when it is in the faddish contagion stage.

Based on the above argument, we propose the following:

Proposition 1: Management fads and fashions has impact on behavioral intention of adopting multi-cloud computing.

At the very initial stage of multi-cloud adoption, organizations first think what IT resources they have to serve their IT needs. Organizations, in this stage, assess their IT capability comparing needed and available IT resources to implement cloud computing (Cheon et al., 1995). Based on the assessment, organizations intend or not intend to outsource cloud computing (Fink, 2010). For example, if organization has enough IT capability, technical assets and human skills, to provide cloud services, it can rely on its own cloud computing architecture. Otherwise, they can rely on their own cloud computing resources for critical activities and can outsource cloud services from vendors for non-critical activities. Therefore, we can argue that IT capability has influence on the decision of multi-cloud adoption. The RBV theory explains that IT capability creates business value for organizations (Fink, 2010). This IT capability consists of technical and human assets. When a firm has superior technical and human assets, it builds its own capability to serve IT needs. In this case, firms are reluctant to outsource multi-cloud adoption. Many studies showed that inferior IT capability leads to rely on outsourcing IT activities (Cheon et al., 1995; DiRomauldo & Gurbaxani, 1998). The TOE framework also acknowledges that a firm that has higher IT capability, it increases the likelihood of adopting new technology (Kamal, 2006; Kuan & Chau, 2001). However, many studies applying TOE framework believed that the inferior IT capability is more likely to adopt cloud computing (Hofmann & Woods, 2010; Sultan, 2011). Based on the above discussion, we propose that:

Proposition 2: IT capability is negatively related to BI of multi-cloud adoption.

When a firm's assessment of IT capability supports to implement multi-cloud adoption, it then assesses the relational capability of adopting multi-cloud computing (Fink, 2010). The "relational view of resources-based theory" extended the core tenets of RBT by integrating the perspective of RBT and relational network theory to explain how inter-firm cooperation can generate mutual benefits (Duschek, 2004; Dyer & Singh, 1998). Dyer and Singh (1998) defined relational capability as ''a supernormal profit jointly generated in an exchange relationship that cannot be generated by either firm in isolation" (p. 662). Relational view also argues that the resources generating competitive advantage can span firm boundaries and embedded in inter-firm relations. Therefore, the sources of competitive advantages are not only from the internal resources owned by a firm itself but also from the external resources in the relational networks. With this line of reasoning, we argue that the relational capability of cloud computing network has positive influence on adoption of multi-cloud computing. Dyer and Singh (1998) stated that the outcome of inter-firm relationships is function of relation-specific assets, knowledge sharing routines, complementary resources/capabilities and effective governance. Fink (2010) pointed out that higher the relational capability between firms, higher the likelihood of being successful of adopting outsourcing. Therefore, we propose that

Proposition 3: The relational capability is positively related the behavioral intention of adopting multi-cloud computing.

The dominant innovation paradigm assumes that organizations are well-armed to implement a new technology that is perceived to beneficial from rational perspective (Fichman, 2004). The theory of management fads and fashions highlighted that sometimes inefficient innovations can be accepted widely, at least for a time. Initially, an innovation may be perceived beneficial because of bandwagon effects. As time goes, the firm realizes that the innovation is rejected in the long run. Therefore, we need to know that when an innovation does and does not conform to the rational ideal, and thereby identify the antecedents and consequences of such conformance (or lack of conformance). To support this task, researchers have begun to consider the concept of mindfulness (Swanson & Ramiller, 2004) and the role it plays in determining organizational innovation, particularly in the context of technologies subject to bandwagon dynamics. Swanson and Ramiller (2004) defined technology mindfulness as a watchful and vigilant state of mind that can distinguish efficient and inefficient innovations. Fiol and O'Connor (2003) argued that an organization that is technologically mindful is most likely going to reject an innovation of dubious merits. Therefore, based on above arguments, we propose the following:

Proposition 4: The relationship between management fads and fashions and behavioral intention of multi-cloud adoption is moderated by technology mindfulness.

In the cloud computing literature, many researchers are concerned with security issues of cloud computing service providers (Ardagna, 2015; Bohli et al., 2013; Hofmann & Woods, 2010; Sengupta et al., 2011). Cloud computing offers many advantages in terms of cost and data usability. Organizations are relying on providers for cloud services storing confidential information, but these vendors can be untrustworthy (AlZain et al., 2012). Therefore, the cloud computing adopting firms are facing challenges of ensuring the security of cloud infrastructure. In this point, we argue that if a firm trusts cloud service providers, it is most likely that the firm will adopt the cloud services from providers. Otherwise, firm may be reluctant to use the cloud services from providers. Many studies found that trust influences the behavioral intention of adopting technology (Bahmanziari et al., 2003; Dahlberg et al., 2003; Srivastava et al., 2010). For example, Dahlberg et al. (2003) found that trust is a significant predictor for explaining tourist technology adoption behavior. On the other hand, trust also plays key role in outsourcing activities (Lee et al., 2008; Lee & Choi, 2011). Since cloud providers are an internet-based organization, data sharing over the Internet is based mostly on “trust” between the service customer and the provider. Rahi, Bisui, and Misra (2017) empirically found that trust moderates the relationships between TOE components and cloud service adoption. Therefore, based on above discussion, we propose:

Proposition 5: Trust positively influences behavioral intention of multi-cloud adoption.

Proposition 6: Trust moderates the relationship between relational capability of adopting firm and providers and behavioral intention of adopting multi cloud adoption.

Figure 1 presents the conceptual model of this paper and propositions.

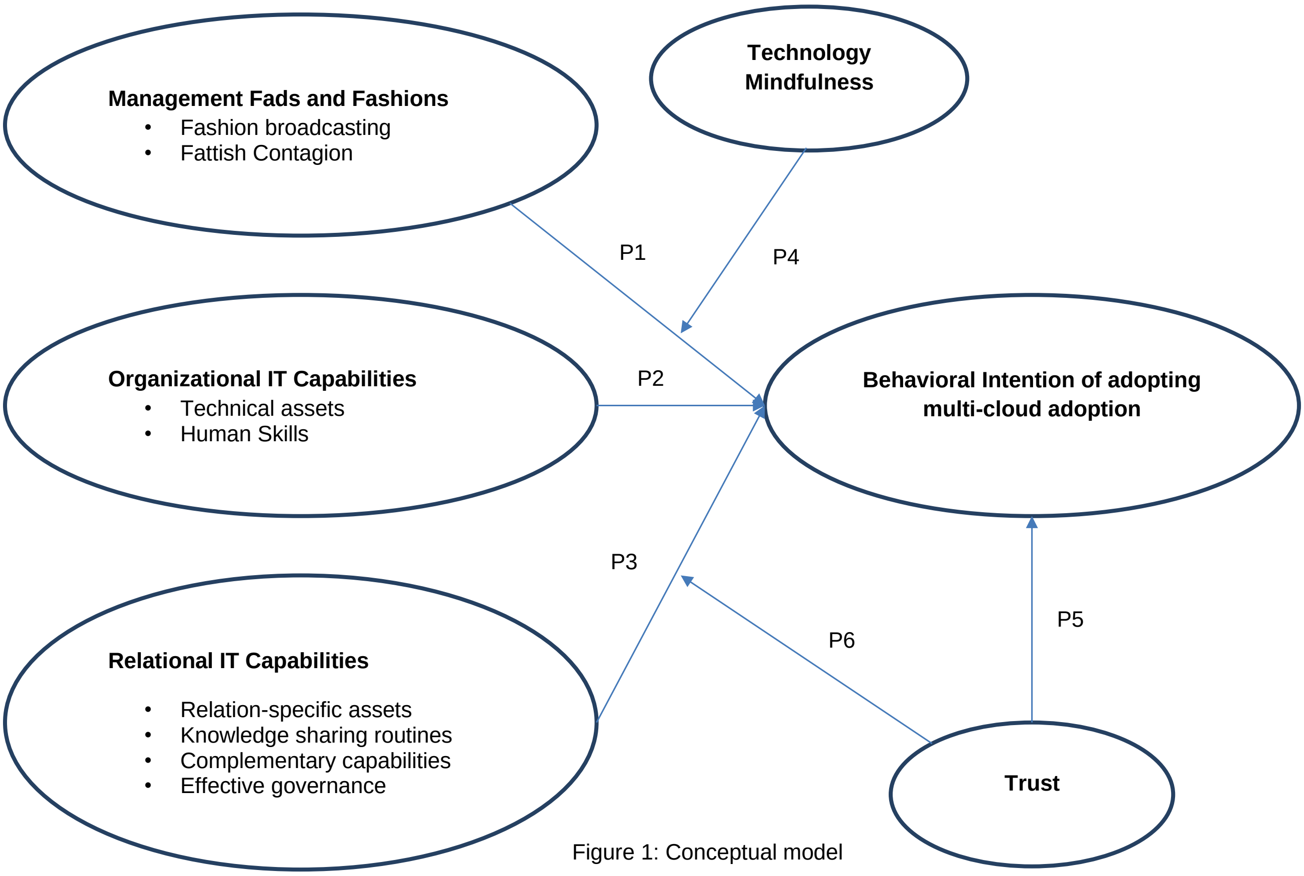


Figure 1: Conceptual model

DISCUSSIONS

Firms are facing problems of adopting and implementing multi-cloud computing because of its convoluted and unclear nature. This multi-cloud computing technology has many aspects that need to be considered and examined before adopting the technology. As this is a new technology, the ultimate outcomes and implementing complexities are not known; therefore, organizations that want to adopt may fall prey of management fads and fashions if it is not mindful. In addition, the use of this technology involves interactions between parties, adopting firm and provider. Hence, the successfulness of the adoption of multi-cloud computing also depends on the relational capability of these parties. Therefore, a critical examination is required to investigate the factors that influence the adoption decision. Moreover, multi-cloud computing is a new innovation that requires big investment and therefore, if not managed properly, it may provide disadvantages leading the firms at stake. Hence, it is mandatory to know how to facilitate and promote the adoption of multi-cloud computing in organizations. Therefore, this study aims to develop a conceptual model of adopting multi-cloud computing. The paper explains the factors that might have influence on the adoption of multi-cloud computing based on different theoretical perspectives. The emerging proposition can be very powerful to explain how and why multi-cloud computing is adopted.

When a new innovation is adopted by a firm in an industry, the other firms in the industry may perceive threats of losing market share. Moreover, industry gurus, e.g., academicians, entrepreneurs and practitioners, may influence firms to adopt the innovation, without knowing the long-term benefits of the innovation. In such case, the innovation adopting firm may be ended up with rejecting the innovation as it may not have merits in future. However, if a firm is mindful, it may distinguish the inefficient innovations.

Overall, this paper proposed 5 dimensions from based on extensive literature review. These dimensions can be used comprehensively as a powerful mechanism to expedite the multi-cloud adoption process. However, these dimensions are very basic and therefore, need more refinement in context of multi-cloud computing. Scrutinizing these dimensions critically may help practitioners adopt suitable strategies that support successful adoption of multi-cloud computing.

RESEARCH IMPLICATIONS

Being an emerging technology, multi-cloud computing has tremendous importance for both practitioners and academicians. To understand the factors that can promote or hinder the use of multi-cloud computing in organizations, theory-based analysis is needed. In this study, we critically reviewed prominent theories to explain how multi-cloud computing is adopted in organizations.

Implications for practitioners

The developed model is very useful for practitioners who want to know how adoption of multi-cloud computing is diffused in a firm. Practitioners should focus on management contagion effects so that this effect cannot play significant role in adopting decision. In addition, managers also need to know what organizational capabilities they have and based on that they should take decision whether multi-cloud computing is going to adopted or not. Our literature shows that security is a major concern for multi-cloud computing. Therefore, managers should pay attention to the trustworthiness of providers, as sensitive information may be stored in provider’s data centers. Finally, managers should be technologically mindful so that they can distinguish inefficient innovations.

Implications for researchers

This study contributes to the existing body of knowledge by providing and explaining underlying constructs that work as determinants of adoption of similar technologies as multi-cloud computing. Moreover, the proposed research model presents the causal relationships between constructs focusing on multi-cloud computing, which will inspire other researchers to examine the phenomena in details synthesizing other relevant information systems theories. In addition, as this study is conceptual, future empirical studies can be conducted to validate the causal relationships. Moreover, future studies can develop and validate the measurements of suggested dimensions. Overall, researchers will have better opportunities to develop and understand related topics surrounding emerging technologies including multi-cloud computing based on this study.

CONCLUSION

Many studies have been undertaken in recent years on various technical innovations such as ERP, but relatively little research has been conducted to ascertain the determinants of multi-cloud computing adoption. In this study, we have developed a conceptual model of multi-cloud computing adoption, which has equal importance for both practitioners and researchers. The suggested model would assist managers in determining what institutional factors influence the multi-cloud computing acceptance and evolution. Finally, future research is directed to examine and validate the proposed relationships empirically.

REFERENCES

Abrahamson, E. (1991). Managerial fads and fashions: The diffusion and rejection of innovations. *Academy of Management Review*, *16*(3), 586–612.

AlZain, M. A., Pardede, E., Soh, B., & Thom, J. A. (2012). Cloud computing security: from single to multi-clouds. *2012 45th Hawaii International Conference on System Sciences*, 5490–5499.

Ardagna, D. (2015). Cloud and multi-cloud computing: current challenges and future applications. *2015 IEEE/ACM 7th International Workshop on Principles of Engineering Service-Oriented and Cloud Systems*, 1–2.

Bahmanziari, T., Pearson, J. M., & Crosby, L. (2003). Is Trust Important in Technology Adoption? A Policy Capturing Approach. *Journal of Computer Information Systems*, *43*(4), 46–54. https://doi.org/10.1080/08874417.2003.11647533

Bohli, J.-M., Gruschka, N., Jensen, M., Iacono, L. Lo, & Marnau, N. (2013). Security and privacy-enhancing multicloud architectures. *IEEE Transactions on Dependable and Secure Computing*, *10*(4), 212–224.

Chang, Y., Lee, H., Lee, J.-N., & Wang, S. (2018). *The adoption and resistance of disruptive information technologies*. EMERALD GROUP PUBLISHING LTD HOWARD HOUSE, WAGON LANE, BINGLEY BD16 1WA, W ….

Cheon, M. J., Grover, V., & Teng, J. T. C. (1995). Theoretical perspectives on the outsourcing of information systems. *Journal of Information Technology*, *10*(4), 209–219.

Dahlberg, T., Mallat, N., & Öörni, A. (2003). Trust enhanced technology acceptance modelconsumer acceptance of mobile payment solutions: Tentative evidence. *Stockholm*

Mobility Roundtable, *22*(1), 145.

DiRomauldo, A., & Gurbaxani, V. (1998). *Strategic intent for IT outsourcing*.

Duschek, S. (2004). Inter-firm resources and sustained competitive advantage. *Management Revue*, 53–73.

Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management Review*, *23*(4), 660–679.

Fichman, R. G. (2004). Going beyond the dominant paradigm for information technology innovation research: Emerging concepts and methods. *Journal of the Association for Information Systems*, *5*(8), 11.

Fink, L. (2010). Information technology outsourcing through a configurational lens. *The Journal of Strategic Information Systems*, *19*(2), 124–141.

Fiol, C. M., & O'Connor, E. J. (2003). Waking up! Mindfulness in the face of bandwagons. *Academy of Management Review*, *28*(1), 54–70.

Frank, B. H., & Jones, S. (2019). *Multi-Cloud Adoption Accelerates*.

Gundu, S. R., Panem, C. A., & Thimmapuram, A. (2020). Hybrid IT and Multi Cloud an Emerging Trend and Improved Performance in Cloud Computing. *SN Computer Science*, *1*(5), 1–6.

Hamel, G., & Prahalad, C. K. (1994). Competing for the future. *Harvard Business Review*, *72*(4), 122–128.

He, W., & Wang, F.-K. (2015). A hybrid cloud model for cloud adoption by multinational enterprises. *Journal of Global Information Management (JGIM)*, *23*(1), 1–23.

Hofmann, P., & Woods, D. (2010). Cloud Computing: The Limits of Public Clouds for Business Applications. *IEEE Internet Computing*, *14*(6), 90–93. https://doi.org/10.1109/MIC.2010.136

Hong, J., Dreibholz, T., Schenkel, J. A., & Hu, J. A. (2019). An Overview of multi-cloud computing. *Workshops of the International Conference on Advanced Information Networking and Applications*, 1055–1068.

Kamal, M. M. (2006). IT innovation adoption in the government sector: identifying the critical success factors. *Journal of Enterprise Information Management*.

Khajeh-Hosseini, A., Greenwood, D., Smith, J. W., & Sommerville, I. (2012). The cloud adoption toolkit: supporting cloud adoption decisions in the enterprise. *Software: Practice and Experience*, *42*(4), 447–465.

Kuan, K. K. Y., & 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. https://doi.org/https://doi.org/10.1016/S0378-7206(01)00073-8

Lahmar, F., & Mezni, H. (2018). Multicloud service composition: a survey of current approaches and issues. *Journal of Software: Evolution and Process*, *30*(10), e1947.

Lee, J.-N., & Choi, B. (2011). Effects of initial and ongoing trust in IT outsourcing: A bilateral perspective. *Information & Management*, *48*(2), 96–105. https://doi.org/https://doi.org/10.1016/j.im.2011.02.001

Lee, J.-N., Huynh, M. Q., & Hirschheim, R. (2008). An integrative model of trust on IT outsourcing: Examining a bilateral perspective. *Information Systems Frontiers*, *10*(2), 145–163.

Petcu, D. (2013). Multi-cloud: expectations and current approaches. *Proceedings of the 2013 International Workshop on Multi-Cloud Applications and Federated Clouds*, 1–6.

Piazza, A., & Abrahamson, E. (2020). Fads and fashions in management practices: Taking stock and looking forward. *International Journal of Management Reviews*, *22*(3), 264–286.

Rahi, S. B., Bisui, S., & Misra, S. C. (2017). Identifying the moderating effect of trust on the adoption of cloud-based services. *International Journal of Communication Systems*, *30*(11), e3253.

Rogers, E. (1983), Diffusion of Innovations, (3rd ed.) New York: The Free Press.

Sengupta, S., Kaulgud, V., & Sharma, V. S. (2011). Cloud computing security--trends and research directions. *2011 IEEE World Congress on Services*, 524–531.

Srivastava, S. C., Chandra, S., & Theng, Y.-L. (2010). Evaluating the role of trust in consumer adoption of mobile payment systems: An empirical analysis. *Communications of the Association for Information Systems*, *27*, 561–588.

Sultan, N. A. (2011). Reaching for the “cloud”: How SMEs can manage. *International Journal of Information Management*, *31*(3), 272–278. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2010.08.001

Sun, Y., & Jeyaraj, A. (2013). Information technology adoption and continuance: A longitudinal study of individuals’ behavioral intentions. *Information & Management*, *50*(7), 457–465. https://doi.org/https://doi.org/10.1016/j.im.2013.07.005

Swanson, E. B., & Ramiller, N. C. (2004). Innovating mindfully with information technology. *MIS Quarterly*, 553–583.

Thong, J. Y. L. (1999). An integrated model of information systems adoption in small businesses. *Journal of Management Information Systems*, *15*(4), 187–214.

DECISION SCIENCES INSTITUTE

Determinants of Cloud Computing Adoption

Charles Changchit
Texas A&M University-Corpus Christi
Email: charles.changchit@gmail.com

Chuleeporn Changchit
Texas A&M University-Corpus Christi
Email: chuleeporn.changchit@tamucc.edu

Robert Cutshall
Texas A&M University-Corpus Christi
Email: robert.cutshall@tamucc.edu

ABSTRACT

Cloud computing is recognized as an important area for IT innovation and investment. Cloud computing is cost effective and sometimes can even be free, depending on the services needed. The costs of IT are also reduced because there is no installing, maintaining, or updating software as well as the elimination of IT emergencies or working after hours. Despite the potential for positive change, there are still other factors to consider before cloud computing can be accepted entirely by the individual, large business, or academic groups. Therefore, the main objective of this study is to examine how individuals perceive cloud computing. The study findings reveal several factors which play an important role in their decision whether to adopt cloud computing technology.

KEYWORDS: Cloud Computing, Technology Acceptance Model (TAM), Perceived Usefulness, Perceived Ease of Use, Perceived Security, Perceived Speed, Perceived Cost, Normative Beliefs, Technology Competency

INTRODUCTION

Cloud computing has had a dramatic impact on the way businesses handle IT (Fan et al., 2015). Cloud computing has spurred the interest of businesses and researchers alike with its ability to revolutionize traditional IT delivery with reduced costs, greater elasticity, and ubiquitous access (Hsu et al., 2014). According to a recent worldwide survey in 2020, 49 percent of respondents state that they plan to deploy a hybrid cloud model within three to five years (Statista, 2021).

The cloud computing model has been increasing in its popularity due to its various benefits such as cost-effectiveness, scalability, usefulness, ease of use, and its anywhere, anytime accessibility. Since 2010, the year-over-year global expenditures on cloud services has increased to $370 billion in 2020, this represents a growth rate of over 380 percent in ten years (Ikink, 2021). The adoption of cloud computing is rapidly increasing, with enterprise applications migrating to the public cloud and organizations becoming more cloud-native in their deployments (Costello, 2021).

It is clear that cloud computing provides many benefits and the potential for positive change; nevertheless, there are other factors to consider before cloud computing can be accepted by individuals, large businesses, or academic groups. Some consumers are still hesitant to adopt cloud computing. Some factors that must be investigated include the amount of file storage, file synchronization, the ability to create documents, as well as how collaboration is achieved (Aaron & Roche, 2012). Additionally, other concerns involve the issues of security, the availability of privacy services, and how the external server will function (Khorshed et al., 2012). Consumers, understandably, want assurance that the documents they store, and share in the cloud will not be accessible to parties outside of their intended readers. Without this guarantee, or if the consumer is not satisfied by the security promises made by the cloud provider, the convenience offered by cloud computing is meaningless if consumers do not have faith in the security of their computing resources. This study thus aims at investigating which factors have an impact on individuals' decision to adopt cloud computing.

LITERATURE REVIEW

Cloud computing is defined by the National Institute of Standards and Technology (NIST) as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Santalesa, 2011). Many new technological innovations such as cloud computing can be attributed to the increasing demand for more advanced and efficient technology. The popularization of cloud computing by companies like Amazon®, Google®, and Apple® ensure that the usage of the cloud as a storage medium for music, movies, and other media content files, will be ubiquitous in the next 10 years (Changchit, 2014).

A previous study by Rightscale (2014) identified crucial factors that could impact the usage of cloud computing such as security, migration, and the availability of internal resources to support its usage. Behrend et al., (2011) led another study examining the factors that could influence the adoption of cloud computing as a virtual computing lab for a class. The findings from that research revealed that subjects' ease of use perception would positively affect intentions for future use, but not for actual use.

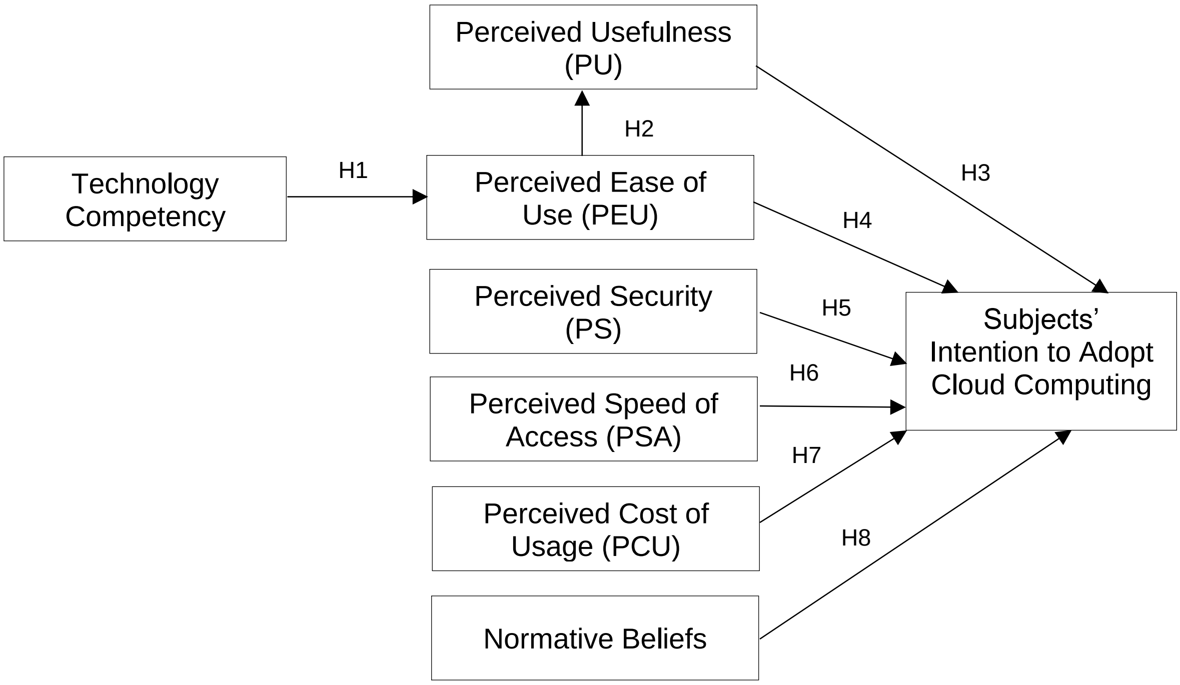
Another study by Fan et al., (2015) was conducted to investigate how most firms react conservatively to adopting cloud computing at the organizational level. The findings confirmed the effect of status quo bias on adoption and revealed that institutional pressures, switching benefits, and switching costs are all salient factors of perceived value (Fan et al., 2015). Gangwar, Date, and Ramaswamy (2015) proposed a model to integrate Technology Acceptance Model (TAM) and Technology Organizational Environment (TOE) framework for cloud computing adoption at the organizational level. The study identified relative advantage, compatibility, complexity, organizational readiness, top management commitment, and training and education as significant variables for affecting cloud computing adoption using perceived ease of use (PEU) and perceived usefulness (PU) as mediating variables. Additionally, competitive pressure and trading partner support were found to directly affect cloud computing adoption intentions. The model explained 62 percent of cloud computing adoption at the organizational level.

Several articles have been written to detail the many benefits that cloud computing can offer to users; however, not everyone is convinced nor are they ready to embrace this technology (Chinyao et al., 2011; Goscinski & Brock, 2010). Before implementing this technology, it is crucial to understand which factors can encourage or discourage users from adopting cloud computing. The findings in this study should help cloud computing service providers, or any organization planning to adopt this technology, understand which factors encourage or discourage their potential users to accept cloud computing, thus increasing the chance for the successful rendering of cloud computing services or the integration of cloud computing into an organization.

THEORETICAL DEVELOPMENT AND RESEARCH MODEL

The Technology Acceptance Model (TAM) was used as a framework for this study. TAM was derived from the Theory of Reasoned Action and developed by Davis in 1986 to introduce the factors impacting users' acceptance of information technology. TAM has seen successful and widespread usage in many different disciplines aside from that of IS. This study extends the TAM to include five additional factors predicted to play a significant role in motivating subjects' to adopt cloud computing. These factors are: (1) Technology Competency (TC), (2) Perceived Security (PS), (3) Perceived Speed of Access (PSA), (4) Perceived Cost of Usage (PCU) and (5) Normative Beliefs (NB). The proposed research model is shown in Figure 1 below.

Figure 1 Research Model



Technology Competency

Technology Competency is defined as an individual's assessment of his or her capability to use computers in a variety of situations (Hsia et al., 2014). People with greater technology competency tend to seek out various technologies to use to accomplish their tasks since they have a higher level of confidence in their ability to accomplish the tasks on a computer. We therefore propose that:

H1: Technology Competency positively affects subjects' intention to adopt cloud computing.

Perceived Ease of Use of Cloud Computing (PEU)

Perceived ease of use plays a critical role in understanding individual responses to information technology (Chau & Hu, 2001). This factor is defined as the degree to which the prospective user expects the target system to be free of effort (Davis et al., 1989). Based on the TAM, this factor has been shown to have an impact on perceived usefulness (Davis et al., 1989). In addition, previous research over the years provides evidence that this factor also has a significant impact on usage intention (Venkatesh & Davis, 2000). We thus hypothesize that:

H2: Perceived ease of use (PEU) positively affects subjects' perceived usefulness.

H3: Perceived ease of use (PEU) positively affects subjects' intention to adopt cloud computing.

Perceived Usefulness of Cloud Computing (PU)

Davis et al (1989) have defined perceived usefulness as the prospective users' subjective probability that using a specific information system will increase their job performance within an organizational context. This factor has a significant impact on usage intention (Davis et al., 1989; Venkatesh & Davis, 2000). A prior study also reported that this construct positively influenced subjects' intention to adopt cloud computing (Changchit & Chuchuen, 2018). Based on the foregoing, we propose the following hypothesis:

H4: Perceived usefulness (PU) positively affects subjects' intention to adopt cloud computing.

Perceived Security of Cloud Computing (PS)

Flavián and Guinalíu, (2006) have defined perceived security as a subjective probability with which consumers believe that their personal information will not be viewed, stored, and manipulated during transit and storage by inappropriate parties in a manner that is inconsistent with their expectations. Since security is always a major concern for all individuals who are dealing with sensitive data in everyday life (Changchit, 2008), there is a high likelihood that subjects should be more willing to adopt cloud computing if they find this technology to be secure. We therefore posit that:

H5: Perceived security (PS) positively affects subjects' intention to adopt cloud computing.

Perceived Speed of Access of Cloud Computing (PSA)

Access to computing resources over the Internet at a reasonable speed is of the utmost importance for cloud computing to be widely accepted (Changchit & Chuchuen, 2018). The speed of using applications over the Internet can be a factor that has prevented cloud computing from being a viable option for outsourcing IT operations. Users may be unaware that the use of applications via the Internet still allows them to retrieve the data at the same speed as when the data is stored locally. Subjects' perception on the speed of access should influence their intention to adopt cloud computing. Hence, we posit that:

H6: Perceived speed of access (PSA) positively affects subjects' intention to adopt cloud computing.

Perceived Cost of Usage of Cloud Computing (PCU)

In 2011, the average cloud expenditure of a typical company was roughly $6,500 per year. In 2020, this number had risen to $10,000 per month (Ikink, 2021). A survey conducted by ComputerWorld magazine (Wood, 2011) of IT professionals revealed that while “saves money” ranked first on the list of cloud computing key benefits, “costs more” ranked third on the list of disadvantages suggesting that the issue of cost in outsourcing information technology resources is a complex issue. In many public and private sector industries, including education, federal and state government, and telecommunications, cloud computing systems are being pilot-tested and implemented to save IT costs and improve performance (Behrend et al., 2011). Subjects perception on the cost of cloud computing should impact their willingness to adopt this technology. We therefore hypothesize that:

H7: Perceived cost of usage (PCU) positively affects subjects’ intention to adopt cloud computing.

Normative Beliefs (NB)

Fishbein and Ajzen (1975) have defined normative beliefs as individuals’ perception that most people who are important to them think they should or should not perform the behavior in question. Limayem et al. (2000) have shown that normative beliefs are significant in the adoption of new technology. Thus, we hypothesize:

H8: Normative beliefs (NB) positively affects subjects’ intention to adopt cloud computing.

RESEARCH METHODOLOGY

Development of Measurement Instrument

The measurement scales for this study utilized the TAM survey instrument and modified it with the addition of five constructs. The perceive ease of use and perceived usefulness factor survey items were revised from the work of Venkatesh & Davis (2000) and Venkatesh et al. (2003). Other survey items to measure the additional factors in this study were developed to identify the significant constructs leading to consumers’ intention to adopt cloud computing. Several tests such as reliability, KMO and Bartlett’s, common method bias, and factor analysis were conducted on this study’s factors to verify and validate their suitability for the measurement model in this study. These results are described in the data analysis section of this paper.

The questionnaire consists of forty-seven (47) questions. Forty questions with the five-point Likert scale were designed to measure subjects’ perceptions on cloud computing and their intention to adopt cloud computing. The remaining seven questions were asked to gather demographic data on the subjects. To validate the clarity of these questions, three professors and three researchers were asked to read through the survey questions. Revisions to the survey were made based on the feedback received.

Data Collection

The surveys were administered to students at a university located in the Southern United States. These subjects as future professionals are certainly part of the target group for companies providing cloud computing services. Four hundred and sixty-seven (467) subjects participated in this study. However, only four hundred and thirty-five (435) responses were usable. Details on the subjects' demographics are provided in Table 1 below.

Table 1: Subjects' demographics (n=435)

	No.	%		No.	%
Gender			**Ethnicity**		
Male	170	39.1	African	5	1.2
Female	223	51.3	Anglo	54	12.4
No answer	42	9.6	Asian	17	3.9
			Hispanic	92	21.2
Age			Native American	3	0.7
18 - 24	347	79.7	Other	0	0.0
25 - 34	48	11.0	No Answer	264	60.6
35 - 44	9	2.1			
45 and over	6	1.4	**Classification**		
No answer	25	5.8	Freshman	77	17.7
			Sophomore	136	31.3
First Generation College Student			Junior	115	26.4
Yes	158	36.3	Senior	41	9.4
No	217	49.9	Graduate	23	5.3
No answer	60	13.8	Other	0	0.0
			No answer	43	9.9
College					
Business	217	49.9	**Employment Status**		
Education	34	7.8	Full-time	45	10.3
Liberal Art	56	12.9	Part-Time	196	45.1
Nursing	38	8.7	Un-employed	151	34.7
Science & Technology	43	9.9	No Answer	43	9.9
No answer	47	10.8			

DATA ANALYSIS AND DISCUSSION

The statistical software package SPSS 26 with AMOS 25 was used to analyze the respondents' data. For SEM, Hair et al. (2009) recommends a sample size of at least 200. Therefore, our sample of 435 exceeds the recommended value.

Reliability and Multicollinearity Test

A reliability test was conducted to examine the internal consistency of the research instrument. The test confirmed the reliability with Cronbach's alpha coefficient of 0.966. In addition, since multicollinearity can have harmful effects (Cenfetelli & Bassellier 2009), multicollinearity was assessed for all of the indicators in the research model. The results revealed that the multicollinearity is not an issue with this data set.

KMO and Bartlett's Test

As shown in Table 2 below, the KMO and Bartlett's Test was conducted to assess the degree of unidimensionality of the scales. The test confirmed the sampling adequacy with the value of

0.954. The Bartlett's test of sphericity showed a p-value of 0.000 for both sets of data. Thus, the null hypothesis was rejected regarding no difference between the correlation matrix and the identity matrix.

Table 2. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.954
Bartlett's Test of Sphericity	Approx. Chi-Square	12190.701
	df	561
	Sig.	0.000

Common Method Bias

To ensure that the model is free from common method bias, which is a measurement error that threatens the validity of conclusions drawn from statistical results, the Harman's single factor test was conducted. The Harman's single factor test which is the most widely used in the literature was conducted (Roni, 2014) and was obtained by running an un-rotated, single-factor constraint of factor analysis in SPSS statistics. As shown in Table 3 below, the variance explained by a single factor of 41.498% is less than the recommended 50% cut-off point (Roni, 2014), indicating that the common method bias is not a major concern in this study.

Table 3. Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	15.264	44.893	44.893	15.264	44.893	44.893	4.330	12.735	12.735
2	2.892	8.505	53.397	2.892	8.505	53.397	3.928	11.552	24.288
3	1.631	4.798	58.196	1.631	4.798	58.196	3.879	11.409	35.697
4	1.561	4.591	62.787	1.561	4.591	62.787	3.686	10.841	46.538
5	1.466	4.312	67.099	1.466	4.312	67.099	2.928	8.611	55.150
6	1.274	3.747	70.846	1.274	3.747	70.846	2.889	8.498	63.648
7	1.224	3.601	74.447	1.224	3.601	74.447	2.509	7.381	71.028
8	1.180	3.471	77.918	1.180	3.471	77.918	2.342	6.890	77.918
9	.546	1.606	79.524						
10	.536	1.578	81.102						
11	.502	1.476	82.578						
12	.434	1.277	83.855						
13	.416	1.223	85.078						
14	.386	1.135	86.213						
15	.384	1.130	87.343						
16	.368	1.083	88.426						
17	.339	.998	89.424						
18	.309	.907	90.331						
19	.291	.855	91.187						
20	.275	.808	91.995						
21	.264	.776	92.770						
22	.258	.759	93.530						
23	.246	.723	94.252						

24	.242	.711	94.963
25	.235	.690	95.653
26	.206	.606	96.259
27	.203	.597	96.857
28	.190	.558	97.415
29	.173	.509	97.925
30	.167	.492	98.417
31	.155	.455	98.872
32	.152	.447	99.319
33	.137	.402	99.721
34	.095	.279	100.000

Extraction Method: Principal Component Analysis.

Factor Analysis

Confirmatory factor analysis with varimax rotation was conducted to examine the construct validity and to verify the groupings of the survey items adopted from previous studies. The results of the factor analysis confirm that the forty survey items distributed themselves into eight factors (see Table 4). The survey items which recorded a value below the suggested reliability level of 0.65 (Hair et al. 2009) were removed from the data analysis.

Table 4: Factor Analysis

Constructs	**Component**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Perceive Usefulness (PU1)	.135	.271	.259	.208	.186	**.694**	.091	.151
Perceive Usefulness (PU2)	.138	.160	.166	.120	.072	**.804**	.145	.150
Perceive Usefulness (PU3)	.219	.264	.197	.229	.210	**.699**	.099	.030
Perceive Usefulness (PU4)	.198	.119	.178	.181	.208	**.696**	.050	.095
Perceived Ease of Use (PEU1)	.219	.190	.270	.200	**.718**	.218	.121	.081
Perceived Ease of Use (PEU2)	.274	.208	.191	.202	**.763**	.221	.130	-.004
Perceived Ease of Use (PEU3)	.227	.151	.224	.224	**.762**	.157	.099	.019
Perceived Ease of Use (PEU4)	.171	.191	.033	.121	**.706**	.100	.187	.256
Perceived Security (PS1)	.119	.154	.228	.197	.175	.085	**.793**	.103
Perceived Security (PS2)	.154	.179	.179	.159	.127	.088	**.848**	.113
Perceived Security (PS3)	.161	.176	.096	.272	.127	.152	**.766**	.086
Perceived Speed of Access (PSA1)	.180	**.757**	.199	.187	.141	.217	.114	.123
Perceived Speed of Access (PSA2)	.195	**.765**	.218	.281	.125	.204	.096	-.005
Perceived Speed of Access (PSA3)	.181	**.769**	.168	.158	.207	.104	.186	.115
Perceived Speed of Access (PSA4)	.214	**.769**	.192	.223	.182	.203	.183	.104
Perceived Speed of Access (PSA5)	.188	**.695**	.282	.254	.158	.165	.133	.149
Perceived Cost of Usage (PCU1)	.083	.156	.166	**.737**	.138	.148	.180	.074
Perceived Cost of Usage (PCU2)	.159	.187	.170	**.736**	.163	.182	.156	.103
Perceived Cost of Usage (PCU3)	.151	.276	.141	**.737**	.175	.169	.167	.147
Perceived Cost of Usage (PCU4)	.145	.266	.265	**.687**	.123	.183	.125	.186
Perceived Cost of Usage (PCU5)	.201	.186	.252	**.702**	.190	.109	.174	.183

Normative Beliefs (NB1)	-.013	.204	.180	.238	.120	.088	.003	**.743**
Normative Beliefs (NB2)	-.081	.028	.119	.188	.139	.085	.132	**.834**
Normative Beliefs (NB3)	-.035	.072	.095	.029	-.011	.121	.103	**.804**
Technology Competency (TC1)	**.823**	.206	.170	.104	.142	.145	.043	-.021
Technology Competency (TC2)	**.751**	.160	.120	.123	.183	.090	.175	-.104
Technology Competency (TC3)	**.837**	.169	.137	.152	.200	.105	.087	-.032
Technology Competency (TC4)	**.849**	.129	.215	.142	.119	.175	.089	-.003
Technology Competency (TC5)	**.836**	.129	.221	.107	.146	.148	.119	.036
Intention to Adopt (INT1)	.259	.262	**.747**	.205	.138	.215	.169	.152
Intention to Adopt (INT2)	.206	.224	**.766**	.227	.175	.218	.201	.162
Intention to Adopt (INT3)	.236	.305	**.733**	.191	.185	.200	.078	.052
Intention to Adopt (INT4)	.186	.178	**.748**	.223	.152	.191	.183	.207
Intention to Adopt (INT5)	.265	.187	**.752**	.247	.167	.169	.137	.119

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

The loadings are <0.65 were removed from the data analysis..

Structure Equation Model

SPSS AMOS 25 was utilized to evaluate the proposed research model. To test the overall goodness of fit of the proposed research model, the measures of df/Chi-square, Goodness of fit, Adjusted goodness of fit, Root mean square error of approximation, Comparative fit index, Tucker Lewis index, and Normed fit index were employed. Table 5 reveals that all the goodness of fit indices fall within their acceptable levels (Bentler & Bonett, 1980; Hu & Bentler, 1999; Tucker & Lewis, 1973). This reveals that the proposed research model exhibited a good fit with the data.

Table 5. Fit Indices for the Models

Fit Indices	**Recommended Value**	**Measurement Model**
Chi-square (CMIN)/df	<=3.00	1.282
Goodness-of-fit (GFI)	>=0.90	0.999
Adjusted goodness-of-fit (AGFI)	>=0.80	0.999
Normed fit index (NFI)	>=0.90	0.998
Tucker Lewis Index (TLI)	>=0.90	0.995
Comparative fit index (CFI)	>=0.93	0.999
Root Mean Square Error of Approximation (RMSEA)	<=0.06	0.025

Hypothesis Testing

Properties of the causal paths including standardized path coefficients are presented in Figure 2. The results of hypothesis testing are shown in Table 6.

Figure 2 Structural equation model path analysis

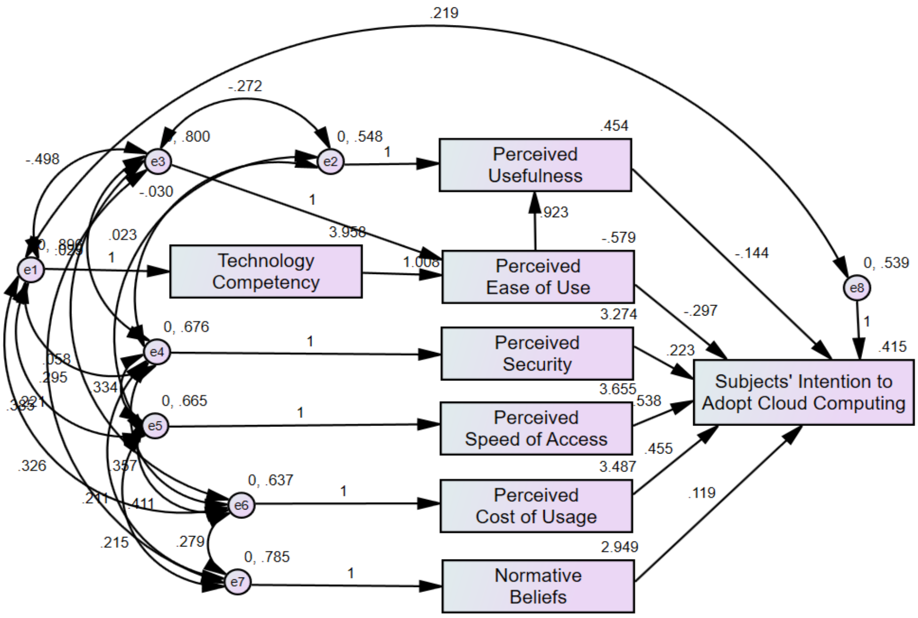


Table 6: Hypothesis testing and results

H#	Hypothesis Testing			Standardized estimate (β)	Critical Ratio	p-value
1	Technology Competency	→	Perceived Ease of Use of Cloud Computing	1.008	11.102	***
2	Perceived Ease of Use of Cloud Computing	→	Perceived Usefulness of Cloud Computing	0.923	14.371	***
3	Perceived Ease of Use of Cloud Computing	→	Intention to Adopt Cloud Computing	-0.297	-2.049	0.040
4	Perceived Usefulness of Cloud Computing	→	Intention to Adopt Cloud Computing	-0.144	-1.048	0.294
5	Perceived Security of Cloud Computing	→	Intention to Adopt Cloud Computing	0.223	3.651	***
6	Perceived Speed of Access of Cloud Computing	→	Intention to Adopt Cloud Computing	0.538	5.109	0.004
7	Perceived Cost of Usage of Cloud Computing	→	Intention to Adopt Cloud Computing	0.455	4.610	***
8	Normative Beliefs	→	Intention to Adopt Cloud Computing	0.119	2.726	0.006

*** indicates significance level < 0.001

For hypothesis H1, the results demonstrate a significant relationship between subjects' technology competency and their perceived ease of use of cloud computing (β = 1.008, p-value < 0.001). This finding indicates that subjects who consider themselves confident with using computers also find it easy to use cloud computing.

The results of hypothesis H2 confirms the relationship between perceived ease of use and perceived usefulness (β = 0.923, p-value < 0.001). The finding is in line with the study conducted by Davis et al. (1989), which indicates that the perceived ease of use can positively influence subjects' perceived usefulness.

Hypothesis H3 examines the relationship between perceived ease of use of cloud computing and subjects' intentions to adopt cloud computing. The results show partial support for H3 by revealing a significant relationship between perceived ease of use and intention to adopt cloud computing (β = -0.297, p-value= 0.040) at the 5 percent level of significance. However, the relationship found was a negative relationship instead of the proposed positive relationship. This significant result confirms the findings of some prior studies (e.g., Koenig-Lewis et al. 2010; Liébana-Cabanillas et al. 2016), but contradicts to other studies (e.g., Dasgupta, et al. 2011; Sripalawat et al. 2011).

Regarding hypothesis H4, the results reveal that the factor perceived usefulness of cloud computing does not have a significant effect on subjects' intentions to adopt cloud computing (β = -0.144, p-value = 0.294). This finding contradicts earlier research by Hanafizadeh et al. (2014) and Püschel et al. (2010) but confirms the findings of research by Lucia-palacios, Pérez-lópez, & Polo-redondo, (2016). It appears that whether subjects believe that cloud computing is useful or not, it does not impact their intention to adopt cloud computing.

Hypothesis H5 reveals the relationship between perceived security of cloud computing and subjects' intention to adopt cloud computing (β = 0.223, p-value < 0.001). This result validates prior studies that perceived security plays an important role in technology acceptance (Alalwan et al. 2017; Bhatt, 2016; Svilar & Zupančič. 2016). It is apparent that subjects are willing to adopt cloud computing if they believe this technology is secure.

For hypothesis H6, the result confirms that there is a significant relationship between perceived speed of access of cloud computing and subjects' intention to adopt cloud computing (β = 0.538, p-value = 0.004). It is pretty obvious that the speed of access should be considered a crucial feature of cloud computing. Subjects who view cloud computing as having a good speed of access are more likely to want to adopt cloud computing.

The results also demonstrate support for H7. The perceived cost of usage of cloud computing is positively related to subjects' intention to adopt cloud computing (β = 0.455, p-value < 0.001). This finding indicates that subjects who perceive the cost of the cloud computing as low are more willing to adopt cloud computing technology.

Lastly, hypothesis H8 reports a positive relationship between normative beliefs and subjects' intention to adopt cloud computing (β = 0.119, p-value = 0.006). This finding conforms to the result in a prior study conducted by Bock et al. (2005). Since subjects usually need to share data among friends and family members, they are more likely to be heavily influenced by the experiences of these people.

CONCLUSION AND STUDY IMPLICATIONS

The results in this study reveal that six out of the seven factors play an important role in encouraging subjects' intention to adopt cloud computing either through an indirect or direct relationship and show partial support for a fifth hypothesis. The results of the data analysis reveal support for hypotheses H1-H2 and H5-H8 (see Table 6). Hypotheses H3 was partially supported by showing significance; however, the relationship was found to be negative. and the hypothesis H4 was not supported. Overall, the independent variables in the research model are able to explain 77.918 percent of the variability in subjects' intention to adopt cloud computing.

The factors which significantly influence the adoption of cloud computing include technology competency, perceived ease of use (partial support), perceived security, perceived speed of access, perceived cost of usage, and normative beliefs. It is very interesting to find that the two original constructs in TAM do not fully impact subjects' intention to adopt the cloud computing technology as anticipated.

The results reveal that regardless of whether subjects perceive cloud computing as a technology that will be beneficial to them or whether they find this technology easy to use, it does not impact their decision to adopt this technology.

The results in this study reveal that subjects with higher technological competency will find the cloud computing easy to use than those who have lower technology competency and thus influence their intention to adopt this technology. In addition, the likelihood that they will adopt cloud computing will be higher if they believe that the technology is secure.

Findings in this study also demonstrate that the speed of access, perceived security, and the cost of using cloud storage are also crucial factors that can influence subjects' intention to adopt cloud computing. Consumers prefer technology that allows them to get what they want in a timely manner. In addition, it does not matter how great a service a new technology can render, it will be useless if they cannot afford the cost of using it.

Based on the findings, normative beliefs also positively influence subjects' intentions to adopt cloud computing. This is quite understandable as cloud computing users will need to share data with their family members and friends. The attitudes of people around them should influence their intention to adopt the cloud computing technology.

As in most empirical studies, there is an inherent limitation in this research. The sample in this research was limited to subjects in one university. Although there was an attempt to gather the data from a variety of courses in the university, future research should be conducted that uses a sample including both students and professionals. Further research should also consider expanding demographics to include users in various countries. In addition, future studies could investigate in more detail which feature of cloud computing will help in the decision to adopt cloud computing.

REFERENCES

Aaron, L. & Roche, C. (2012). Teaching, learning, and collaborating in the cloud: Applications of cloud computing for educators in post-secondary institutions. *Educational Technology Systems*, 40(2), 95-111

Alalwan, A. A., Dwivedi, Y.K., & Rana, N.P. (2017). Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust. *International Journal of Information Management*, 37(3), 99-110.

Behrend, T., Wiebe, E.N., London, J., & Johnson, E. (2011). Cloud computing adoption and usage in community colleges*. Behavior & Information Technology*, 30(2), 231-240.

Bentler, P., & Bonett, D. (1980). Significance tests and goodness-of-fit in analysis of covariance structures. *Psychological Bulletin*, 88(3), 588-606. DOI:10.1037/0033-2909.88.3.588.

Bhatt, A. (2016). Factors affecting customer's adoption of mobile banking services. *Journal of Internet Banking and Commerce*, 21(1), 1-22.

Bock G.W., Zmud, R.W., Kim Y.G., & Lee J.N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly*, 29(1), 87–111.

Cenfetelli, R. T., & G. Bassellier. G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly,* 33(4), 689–707.

Changchit, C. (2008). Data protection and privacy issue. *Journal of Information Privacy and Security*, 9(3), 1-2.

Changchit, C. (2014). Students' perceptions of cloud computing. *Issue in Information System*, 15(1), 312-322.

Changchit C. & Chuchuen, C. (2018). Cloud computing: An examination of factors impacting users' adoption, *Journal of Computer Information Systems*, 58(1), pp. 1-9.

Chau, P.Y.K., & Hu, P.J.H. (2001). Information technology acceptance by individual professionals: a model comparison approach. *Decision Sciences*, 32, 699-719.

Chinyao, L., Yahsueh, C. & Mingchang, W. (2011). Understanding the determinants of cloud computing adoption. *Industrial Management and Data Systems*, 111(7), 1006-1023.

Costello, K. (2021). Gartner Predicts the Future of Cloud and Edge Infrastructure. Retrieved from https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-cloud-and-edge-infrastructure/

Dasgupta, S., Paul, R., & Fuloria, S. (2011). Factors affecting behavioral intentions towards mobile banking usage: Empirical evidence from India. *Romanian Journal of Marketing*, 6(1), 6-28.

Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35, 982-1003.

Fan, Y., Wu, C., Chen, C., & Fang, Y. (2015). The effect of status quo bias on cloud system adoption. *The Journal of Computer Information Systems*, 55(3), 55-64.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley, Reading, MA, USA.

Flavián, C., & Guinalíu, M. (2006). Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. *Industrial Management & Data Systems*, 106, 601-620.

Gangwar, H., Date, H., & Ramaswamy, R. (2015). Understanding determinants of cloud computing adoption using an integrated TAM-TOE model. *Journal of Enterprise Information Management*, 28(1), 107-130.

Goscinski, A., & Brock, M. (2010). Toward dynamic and attribute based publication, discovery and selection for cloud computing. *Future Generation Computer Systems*, 26(7), 947-970

Hair, J., W. Black, B. Babin, & R. Anderson. 2009. *Multivariate data analysis: A global perspective*. 7th ed. Upper Saddle River: Prentice Hall.

Hanafizadeh, P., Keating, B.W., & Khedmatgozar, H.R. (2014). A systematic review of Internet banking adoption. *Telematics and Informatics*, 31(3), 492–510.

Hsia, J.W., Chang, C.C. & Tseng, A.H. (2014). Effects of individuals' locus of control and computer self-efficacy on their e-learning acceptance in high-tech companies. *Behaviour & Information Technology*, 33(1), 51-64

Hsu, P., Ray, S., & Li-Hsieh, Y. (2014). Examining cloud computing adoption intention, pricing mechanism, and deployment model. *International Journal of Information Management*, 34, 474-488.

Hu, L.T., & Bentler, P. M. (1995). Evaluating model fit. In Hoyle R.H. (Ed.), Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage.

Ikink, R. (2021). 25 cloud trends for 2021 and beyond. Retrieved from https://www.accenture.com/nl-en/blogs/insights/cloud-trends

Khorshed, M., Shawkat, A., & Wasimi, S. (2012). A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. *Future Generation Computer Systems*, 28(6), 833–851.

Koenig-Lewis, N., Palmer, A., & Moll, A. (2010). Predicting young consumers' take up of mobile banking services. *International Journal of Bank Marketing*, 28(5), 410-432.

Liébana-Cabanillas, F., Alonso-Dos-Santos, M., Soto-Fuentes, Y., & Valderrama-Palma, V.A. (2016). Unobserved heterogeneity and the importance of customer loyalty in mobile banking. *Technology Analysis & Strategic Management*. DOI: 10.1080/09537325.2016.1262021.

Limayem, M. Khalifa, M., & Frini, A. (2000). What makes consumers buy from Internet? A longitudinal study of on-line shopping, *IEEE Transaction System Man, and Cybernetics*, 30(4), 421–432.

Lucia-palacios, L., Pérez-lópez, R., & Polo-redondo, Y. (2016). Enemies of cloud services usage: Inertia and switching costs. *Service Business*, 10(2), 447-467. doi:http://dx.doi.org/10.1007/s11628-015-0277-y

Püschel, J., Mazzon, J.A., & Hernandez, J.M.C. (2010). Mobile banking: Proposition of an integrated adoption intention framework. *International Journal of Bank Marketing*, 28(5), 389-409.

Rightscale. (2014). 2014 State of the cloud report: see the latest trends on cloud adoption. Retrieved from http://www.rightscale.com/lp/2014-state-of-the-cloud-report.

Roni S.M. (2014). Introduction to SPSS. Researchgate. 2014 June. Retrieve from http://www.researchgate.net/publication/262151892_Introduction_to_SPSS.

Santalesa, R. (2011). Definition of cloud computing – NIST releases final SP 800-145. Retrieved from http://www.infolawgroup.com/2011/10/articles/cloud-computing1/definition -of-cloud-computing-nist-releases-final-sp-800145/

Sripalawat, J., Thongmak, M., & Ngramyarn, A. (2011). M-banking in metropolitan Bangkok and a comparison with other countries. *Journal of Computer Information System*, 51(3), 67-76.

Statista. (2021). Current and planned cloud deployment models worldwide from 2020 to 2025, by deployment model. Retrieved from https://www.statista.com/statistics/1209565/current-future-it-cloud-models/

Svilar, A., & Zupančič, J. (2016). User experience with security elements in Internet and mobile banking. *Organizacija*, 49(4), 251-260.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10. https://doi.org/10.1007/BF02291170

Venkatesh, V., & Davis, F.D. (2000). "A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science*, 46(2), 186-204.

Venkatesh, V., Morris, M.G., Davis, G.B. & Davis, F.D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27, 425-478.

Wood, L. (2011). Cloud storage a steep climb. *ComputerWorld*, 20-30.

DECISION SCIENCES INSTITUTE

Society and sites of conscience: operational excellence in prison industries

Dan Bumblauskas, Ph.D.
University of Northern Iowa
Email: daniel.bumblauskas@uni.edu

Russell Guay, Ph.D.
University of Northern Iowa
Email: russell.guay@uni.edu

Steve Forbes
Iowa Prison Industries
Email: steve.forbes@iowa.gov

Salil Kalghatgi
Cook Medical
Email: salil.kalghatgi@cookmedical.com

ABSTRACT

Prisons can utilize effective operations management and continuous improvement techniques to fulfil the obligations of stakeholders and provide an opportunity to try and support the rehabilitation of inmates. Iowa Prison Industries (IPI) offers a work training program for incarcerated individuals under the supervision of the Iowa Department of Corrections (IDOC). The program works with approximately 800 incarcerated “associates” to help them acquire skills necessary to obtain gainful employment upon release. This article shares some of the history of IPI’s Operational Excellence (OpEx) as well as reports on program impact. Preliminary findings and success outcomes have been encouraging.

KEYWORDS: Prisons, Operations Management, Six Sigma, Lean, Decision Sciences in Practice

Introduction/Purpose/Motivation

Prisons and correctional facilities across the world service a unique purpose to “correct” legal and societal miscues. Prisons can utilize effective operations management techniques to fulfill the obligation of stakeholders and provide an opportunity to try and support the rehabilitation of their inmates. The framework of human capital theory can be applied to examine the relationships between training programs and gainful future employment for incarcerated individuals (Flatt & Jacobs, 2018). Human capital theory (Schultz, 1961) argues that direct expenditures on education and skills training will all lead to increases in individuals’ income via increased employability and productivity within organizations. This can also lead to “economic mobility” for workers. In the state of Iowa (USA), Iowa Prison Industries (IPI) offers a workforce-development training program for incarcerated men and women under the supervision of the Iowa Department of Corrections (IDOC).

The program works with approximately 800 incarcerated “associates” to help them acquire both job-specific hard and soft skills (Becker, 2007; Flatt & Jacobs, 2018; Valentine & Redcross,

2015) necessary to obtain gainful employment upon release. IPI has what is known as "traditional industries" which are the common training programs many states operate nationwide. They most often include the manufacturing of road signs, license plates, office furniture, graphic art, cleaning chemicals, etc. and recently expanded into COVID-19 related solutions like hand sanitizer and masks. IPI also operates a Private Industry Enterprise (PIE) where they partner with private industries to manufacture goods or services either inside the institutions or on-site at the partner company's location. Both the IPI traditional and PIE programs have rules and guidelines set by the legislature in the Iowa Code around things such as who their customers can be, where their funds arrive from, and who can participate in the different programs.

This article will share some of the history of IPI's operations as well as report on the impact of their programs including a brief history of the IPI work training program and a summary of the different training programs across the state with examples of the hard and soft skills incarcerated men and women can obtain. This article also includes an overview of the certificates and recognized apprenticeship opportunities available to them. Preliminary findings and success outcomes (e.g., recidivism rate) have been encouraging. A meta-analysis of 58 studies from 1980-2011 shows that "inmates who participated in correctional education programs had 43 percent lower odds of recidivating than inmates who did not." (Rand, 2013).

Introduction/Design/Methodology/Approach

Background and history: Iowa Prison Industries

IPI is not appropriated tax dollars and thus must provide high quality goods and services to sell to customers to remain sustainable. An important element of this work is the use of operational excellence methods and tools, such as Lean Six Sigma, to provide a direct impact to IPI and an indirect impact to the 'associates' as they are re-introduced to the civilian labor force outside the prison walls. For IPI's traditional industries, they often only sell goods and services to a limited customer base such as other government agencies including federal, state, county, and city; non-profit groups including churches, fraternal organizations, parochial schools, and school districts; as well as universities, hospitals, and nursing homes.

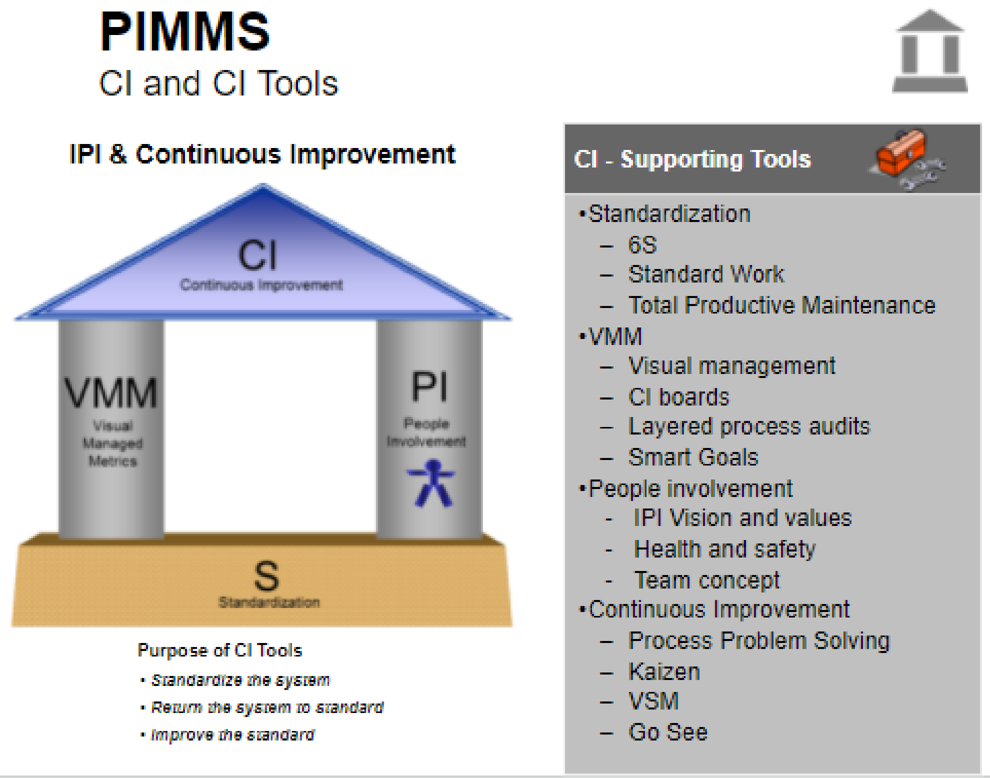
The creation of the Prison Industries, Manufacturing & Management System (PIMMS)

While Prison rehabilitation programs, including undergraduate degree programs such as those at Pitzer College (St. Amour, 2021) are not new, IPI has developed a unique program. In 2015, the leadership of IPI recognized the benefit and need to add a formal lean and continuous improvement program to IPI. Following a process improvement project, it was decided to create a permanent Training Specialist position and focus resources on creating and deploying a lean program structure across the operation.

The initial program that was developed used a two-pillar structure and represents the involvement of every leader, frontline staff member, and incarcerated individual working in the program. The PIMMS program was developed with the understanding that a strong foundation based on standardization would be the supporting base for the other tools and work methods that make up the "PIMMS Structure". Those foundational tools are 6S, Standard Work, and Total Productive Maintenance (TPM). Figure 1 provides an illustration of the PIMMS model and associated continuous improvement tools.

Lean Six Sigma is more than a set of tools and ultimately results in the empowerment of end-users to take ownership of their overall workflow (Bumblauskas & Kalghatgi, 2018). There is a unique need for workforce training in prisons as incarcerated individual may have long gaps in employment and need development to bring additional value when they join a manufacturing work environment outside the prison industries. Through the adoption of PIMMS, a program that advocates Lean Six Sigma, prisoners have a unique ability to be leaders with advanced skills and, hopefully, also feel more empowered in their work and life after prison.

Figure 1 – IPI PIMMS & CI Tools



Of the foundational tools, IPI has discovered that 6S has been the ideal tool to introduce continuous improvement. We have often seen the use of 5S - Sort, Set in order, Shine, Standardize, and Sustain - used in contexts such as public service to help introduce a culture to Lean Six Sigma (Bumblauskas & Kalghatgi, 2018), but have expanded this at IPI to 6S with the addition of 'Safety' to the methodology. Being an organizational tool, it is very visual, generates almost immediate results of improvement, and is easily understood by all. Portions of 6S have been used in prison settings for years without most recognizing it. Shadow boarding tools to assist security for tool control have been in place for decades. Since 2015, IPI has taken 6S to new levels in nearly all of their operations at nine locations across the state. Each of 36 teams have held a 6S "event" that consisted of 6S training, setting the scope including objectives & goals, and then implementing those changes. While 6S is widely used as an organizational tool to improve efficiency, it also has significant security-related benefits in the correctional industries environment. Anything in the area that is not needed is removed during the "sort" process of 6S. Many of those items can easily become contraband or possibly be converted into weapons in the wrong hands. In addition, IPI has also greatly improved the line of sight for staff by eliminating unnecessary materials racking and redesigning cell layouts that eliminate blind spots for staff and surveillance cameras. These improvements have created a much safer work area for everyone. For example, Figure 2 shows before and after photos of 6S improvements in the Anamosa, Iowa Metal Furniture operation.

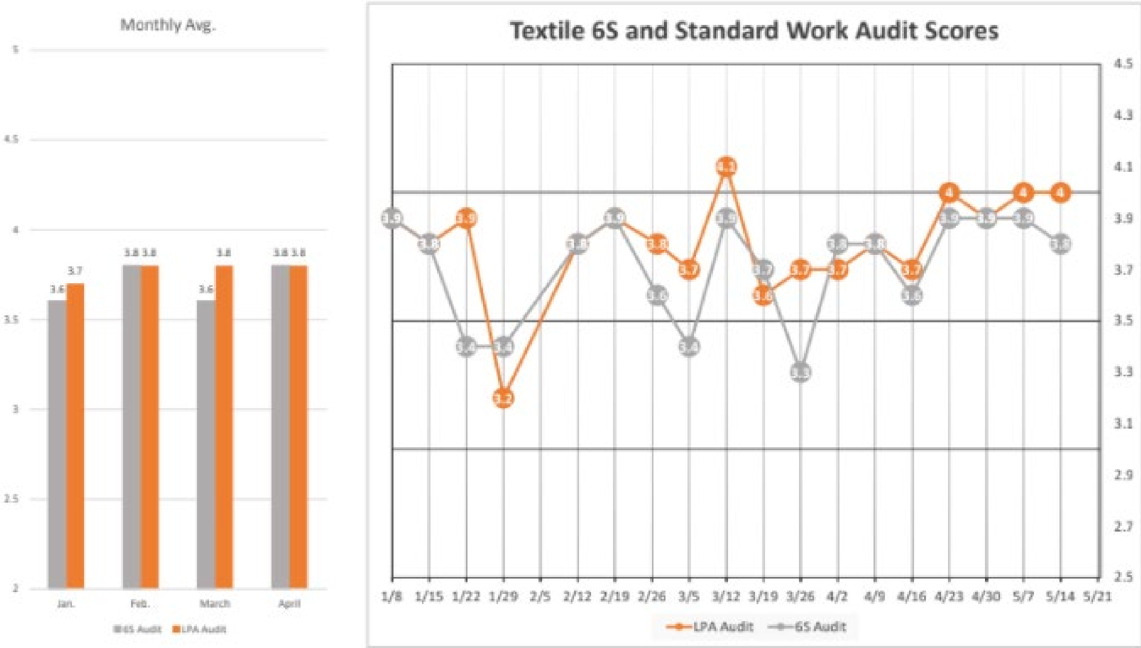
Figure 2 – 6S Before and After

Before:

After:

To sustain the 6S work being done, IPI has put in place a weekly 6S audit process. All of IPI's leadership, staff, and incarcerated individuals have the opportunity to participate in completing weekly 6S audits that provide feedback detailing examples of things going well and opportunities for improvement. The audits use a scoring system that ranges from 1 to 5. Anything that scores a 1 or 2 is an improvement opportunity and gets noted on the teams' metric boards. The item then gets assigned to someone to address. That individual will decide the corrective action, set the time frame, and do any status updates on the metric board so all the team members are informed of the progress. Scoring a 3 is average to good. A score of 4 means something is better than average and it could be an example of a best practice to share with other teams. Scoring a 5 is something exceptional and is potentially worthy of sharing with others to become a new standard for everyone to adopt. The audits are broken up into sections with several questions for each of the 6S categories -- Safety, Sort, Set in order, Shine, Standardize, and Sustain. Figure 3 below is an example of how the Fort Dodge IPI textiles team tracks their 6S and Standard work audit (LPA) performance over time.

Figure 3 – 6S Audits Scoring



Some examples of improvements made due to the 6S portion of PIMMS at IPI include:

- Many areas were cleaned and painted to make it a more pleasant place to work.
- Shop lighting was replaced with brighter and more energy efficient LED lights.
- Labeling of all the equipment and materials was done to make it visual for all.
- Office operations developed a naming system for electronic files so they would be easier for everyone to find.
- Min / max levels setpoints were established and displayed for many raw materials and finished goods.
- Production areas were moved to create better process flow through the areas.
- Additional floorspace was created for new production. In one specific instance, 400 sq ft was freed up and a housekeeping chemical mixing and bottling operation was expanded creating additional jobs.
- A materials storage building was converted into space to do packaging for a private sector operation creating over 100 additional jobs in the Rockwell City location. Similar efforts in Mitchellville, Fort Dodge, and Mount Pleasant have allowed for expanding that partnership which created an additional 200 positions.
- All areas have either painted or had floor tape installed using various colors based on industry standards to indicate where movable items (such as pallet jacks, trash cans, material carts, etc.) should go. They also mark the pedestrian walkways, electrical and safety area clearances, and production equipment. Different colored areas have also been added for raw materials, finished goods, work in process, as well as non-conforming items.

Figure 4 provides an example of the color code system used from a company that supplies tape and labeling materials. That company is Creative Safety Supply.

Figure 4 – Supplier Floor Tape Taxonomy (source: Creative Safety Supply)

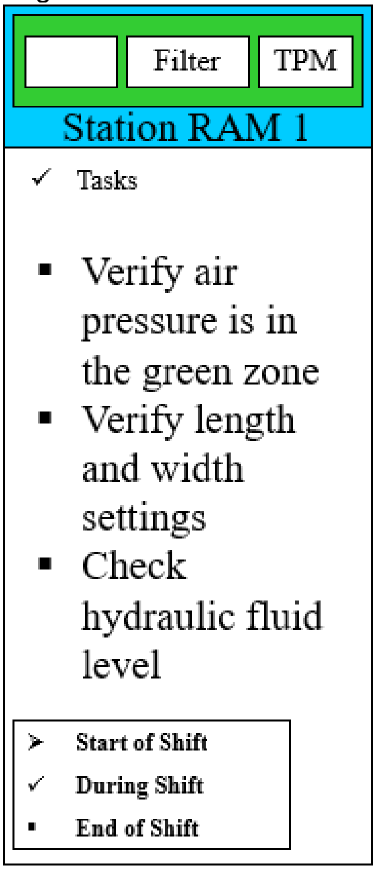
Yellow	Aisleways & Traffic Lanes; Paths of Egress; Work Cells
White	Production; Racks, Machines, Carts, Benches, & Other Equipment
Red	Defect/Scrap Area; Red Tag Area
Orange	Material or Product Inspection; Energized Equipment
Green	Materials & Manufacturing: Finished Goods
Blue	Materials & Manufacturing: Raw Materials
Black	Materials & Manufacturing: Works in Progress
Black & Yellow	Areas which present physical or health risks to employees
Red & White	Areas to be kept clear for safety reasons
Black & White	Areas to be kept clear for operational purposes

The other lean tools that make up the foundation of standardization for PIMMS are standard work documentation and total productive maintenance (TPM). Documented standard work is an especially useful tool in a prison industries type of environment. The goal of any prison industries work program is to teach and develop skills so when those individuals are released, they have a greater opportunity to obtain gainful employment (Becker, 2007; Flatt & Jacobs, 2018; Valentine & Redcross, 2015) so naturally, prison industries work programs want to have significant turnover of individuals. The standard work developed in IPI's program is a collaborative effort by qualified operators that agree upon the safest, most efficient method known. By using this approach, IPI has found operators have much greater ownership and discipline to follow the process as it is written. Following the documented standard work is critical for continuous improvement. Without standards, there can be no continuous improvement. Standard work is the wedge that keeps the Plan-Do-Check-Act process of continuous improvement moving forward. Another unique situation related to correctional industries is the fact that operations can be impacted at a moment's notice due to a security-related incident. Contraband or a fight can put part of or the entire operation into lockdown status. Depending on the situation, operations may be down for long periods of time. Additionally, the current qualified operator may or may not return. While no standard work can capture every detail or replace experience, professionally written standard work provides a starting point for training new operators in the event of unsuspected personnel changes.

Because standard work is an important part of the PIMMS foundation, IPI developed standard work audits for the operators and staff to perform weekly. These audits are interactive. The auditor reinforces that they are auditing a process and not the individual. The audits cover topics such as safety awareness and who to ask when they have questions or need help. When someone does need help, the auditor watches the process, following along with the documented standard work to ensure the steps are being followed in the proper order along with any key points or required details followed. Doing the audits has the added benefit of helping the team discover opportunities to improve the documentation or add visual aids for important steps if the auditor has questions or struggles to understand what the documents are trying to show.

The final tool that comprises the foundation of the PIMMS structure is Total Productive Maintenance or TPM, which is an operator-owned maintenance program that utilizes their skills to keep the machinery in the best working condition possible. The operator that runs the machine daily is in the best position to recognize any changes in performance, leaks, vibrations, or any unfamiliar noises or odors. Those observations will catch any changes early and reduce or eliminate catastrophic machine failures. It also allows the team to plan for preventative maintenance downtime when it fits best into the production schedule. Operators have daily task cards for items they can be trained to check such as machine speed or temperature settings, pressure levels, and any fluid levels. The task cards list the station, the type of task required (6S or TPM), what the task is, and when it should be completed (e.g., start of shift, during shift, or end of shift). All machine control adjustments and any gauges are verified that they are operational, and gauges are color coded. The operators are trained that if the needle is in the green zone, it is good to operate. All fluid reservoirs are marked with the operating range and what type of fluid to use. Any lubrication points are color coded to identify if they need to be greased daily, weekly, monthly, etc. Figure 5 below provides is an example of a TPM task card from the Anamosa Filter shop.

Figure 5 – TPM Task Cards



Another important aspect of TPM being used by IPI is using Overall Equipment Effectiveness (OEE) calculations to find ways to understand capacities and address the variables that make production improvements. OEE is calculated by multiplying the percentages of available minutes times the efficiency rate times the defect rate. Many of the teams were shocked to discover in their normal 8-hour day (480 minutes), a large percentage of that time was not being used for production. Clocking in, checking out tools, daily meetings, two prisoner counts, breaks, lunch, checking tools back in, and clocking out all took away significant production time. The other pieces of the OEE equation look at losses due to machine slowdowns and set-ups as well as the impact of product damage due to tooling or machine operation. By collecting the data for the equation and examining the process, many teams have found ways to improve their output capacities.

The PIMMS structure has two pillars. One pillar focuses on visually managed metrics and operations. The other pillar focuses on people involvement. Visually Managed Metrics is based on all IPI's 36 teams having an aforementioned metric board where each team meets daily. They discuss successes, plans, and opportunities. Each team also sets SMART (specific, measurable, attainable, realistic, and time-bound; Locke & Latham, 1990) goals that their work can influence and align with the overall goals for the entire IPI operation statewide. The areas measured are Safety, Quality, Delivery, and Sustainability. The metric board is designed to be a central hub for information for the team. There is an area for the team's challenge, announcements, their 6S and Standard work audits and performance plotted over time, and a countermeasure sheet where the team records and tracks any opportunities for improvement.

The team meets daily and asks the question," Did we meet our SMART goals for today?" If they did, they color the day in green. However, if they did not, they color that day in red and identify the opportunity on the countermeasure sheet. When they have problems or opportunities, they also have a Process Problem Solving Sheet (PPS) on the board. A PPS is designed to be a team brainstorming tool that examines the possible causes of a process problem using a fishbone diagram, and then uses the 5 why's tool to identify and document the possible root cause of the issue. Figure 6 is an example of the IPI team metric board.

Figure 6 – IPI Metric [Huddle] Board

The second pillar in the PIMMS structure is for People Involvement. The focus on this part of the structure is to teach and develop skills in individuals so they have a greater chance of obtaining a job upon release. The vision of the IPI program is to provide a meaningful work training opportunity to develop utilizable skills for every person. Those skills not only include a variety of work skills, but also the soft skills of working together as a team, maintaining good attendance, and developing problem solving skills. Meeting daily as a team, leading the daily meetings, and performing weekly 6S and standard work audits all help develop these skills. In addition, IPI has a Continuous Improvement Idea program where individuals can submit an idea for improvement and earn a small bonus of cash, treats (soda or candy bars), or a combination of both. There are 3 levels of bonuses which are determined by the amount of effort and thought put into the idea as well as the impact the change has on the operation. There have been ideas ranging from improving safe working conditions to reducing scrap or improving process outputs. There have been thousands of ideas submitted since starting this process in 2018. For those areas that track their idea submissions, it is common to have over 80% of the ideas submitted be implemented. Another powerful tool included in the People Involvement area of PIMMS is the use of quarterly reviews. IPI has tested a pilot group and is in the process of rolling out a program where every individual working in any role at any location will once a quarter have the opportunity to sit down face to face with their supervisor and share feedback in both directions. They will discuss what is going well and an opportunity to further develop some area or skill. The reviews are documented and for any incarcerated individual, those documented reviews will be available to them upon being released so they may share them with future employers if they choose to do so.

Every year, IPI works to find ways to improve their PIMMS program. It was recognized that to make improvements to PIMMS, there needed to be a way to measure or gauge the progress and success of each team deploying PIMMS. The tool that was developed to measure the progress of PIMMS deployment is called the belt level assessment tool. The belt level assessment tool has 5 categories. The foundational tools of PIMMS are each a category. Those include 6S, Standard Work, and TPM. The other two categories include Visual Metrics and Engagement. Desired behaviors were then defined for each category and assigned to a belt color level like those used in martial arts and academic / industry Six Sigma programs.

Figure 7 – IPI Belt Level Assessment

WHITE	Introduction
YELLOW	Awareness
ORANGE	Knowledge
GREEN	Involvement
BLUE	Engagement
BROWN	Ownership

Findings/Relevance/Contributions

Each industry, past and present, gives the chance at a better, brighter future for Iowa's incarcerated individuals. Having meaningful work training opportunities during incarceration has two primary benefits. First, it develops skills in the incarcerated individual that will help them have a new start in their life (Becker, 2007; Flatt & Jacobs, 2018; Valentine & Redcross, 2015). Often, incarcerated individuals have never held a job prior to coming to prison. The skills they can acquire include not only specific trade work skills, but also essential skills such as good attendance, being on time, working well with teams, making good decisions, basic reading & math skills, interpersonal skills, as well as personal characteristics and attitudes (Tonkin, Dickie, Alemagno, & Grove, 2008) are crucial for a returning citizen to possess. These may sound simple, but to an adult individual that may not have a history of traditional employment , these are key . Second, the skills training creates a new level of confidence when reentering the workforce and society (Pearson, Lipton, Cleland, & Yee, 2002).

Conclusion and Future Work

Between 2015 and 2021, IPI has continued to develop PIMMS which uses lean tools and continuous improvement thinking to solve problems and create further opportunities. Developing this mindset and culture has opened the door to expanding the program by more than 200 additional work training opportunities for the men and women incarcerated with the IDOC. Most of these individuals working in the IPI program will be released upon completion of their sentence. By developing work skills along with the problem-solving skills gained within the PIMMS program, the individuals in the IPI program are much more likely to be equipped to find a good paying job which greatly increases their chances of a successful re-entry into the community. While preliminary results on the recidivism rate for all the IDOC vs. those that have participated in the IPI work training program have been promising, more data is being collected to validate this.

References

Becker, G.S. (2007). Health as human capital: Synthesis and extensions. *Oxford Economic Papers, 59,* 379-410.

Bumblauskas, D., & Kalghatgi, S. (2018). Order in the courts: propagating a culture of efficiency within the Irish Court Services. Public Money & Management, 38, 271-280.

Flatt, C., & Jacobs, R.L. (2018). The relationship between participation in different types of training programs and gainful employment for formerly incarcerated individuals. *Human Resource Development Quarterly, 29,* 263-286.

Locke, E.A., & Latham, G.P. (1990). A theory of goal setting and task performance. Upper Saddle River, NJ: Prentice-Hall.

Pearson, F.S., Lipton, D.S., Cleland, C.M., & Yee, D.S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime & Delinquency, 48, 476-496.*

Rand Corporation (2013). Education and Vocational Training in Prisons Reduces Recidivism, Improves Job Outlook, Rand Corporation, August 22, 2013, https://www.bja.gov/Publications/RAND_Correctional-Education-Meta-Analysis.pdf

Schultz, T.W. (1961). Investment in human capital. *The American Economist Review, 51,* 1-17.

St. Amour, M. (2021). B.A. Program for Students, Incarcerated and Not. Inside Higher Ed. Available: https://www.insidehighered.com/news/2021/01/05/pitzer-college-creates-first-ba-program-open-both-its-students-and-incarcerated

Tonkin, P., Dickie, J., Alemagno, S., & Grove, W. (2008). Women in jail: “Soft skills” and barriers to employment. *Journal of Offender Rehabilitation, 38,* 51-71.

Valentine, E.J., & Redcross, C. (2015). Transitional jobs after release from prison: Effects on employment and recidivism. *IZA Journal of Labor Policy, 4,* 16-31.

Further Reading

Dal, B., Tugwell, P., & Greatbanks, R. (2000). Overall equipment effectiveness as a measure of operational improvement–a practical analysis. International Journal of Operations & Production Management.

Ljungberg, Õ. (1998). Measurement of overall equipment effectiveness as a basis for TPM activities. International Journal of Operations & Production Management.

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. International journal of production research, 46(13), 3517-3535.

DECISION SCIENCE INSTITUTE

Population Growth – A Public Health Perspective

Grace L. Edwards
RM 0418 Du Bois
3900 Walnut St. MB
Philadelphia, PA 19104-6138
graceedw@sas.upenn.edu
240 308-2473

Jude E. Edwards
1401 Dorset Lane,
Wynnewood, PA 19096
drjudeedwards@gmail.com
301 814-0345

Layla Halawi
202 Windward Passage,
Unit 308
Clearwater Beach, FL 33767
813 454-3789
Halawil@erau.edu

ABSTRACT

The United Nations classifies countries as "Developed," "Developing," or "Least Developed" based on 1) Human Development Index (HDI); 2) Political Stability; 3) Gross Domestic Product (GDP); 4) Industrialization; 5) Freedom. This explorative study aims to examine population data from developed and underdeveloped countries and cite possible reasons that impact the quality of life in countries. For example, across the U.S., societal factors - education disparities, environmental racism, economic inequality between cities, suburban, and rural areas, contribute to the quality of life. Niger and Norway allow for a comparative examination of factors that contribute to population growth and public health.

KEYWORDS: Economic Inequality, Environmental Racism, Freedom, Human Development Index, Industrialization, Infrastructure, Low Human Development, Maternal Mortality, Population Growth, Political Stability, Poverty, Public Health, Sustainable Development.

INTRODUCTION

As of 2018, Norway has an HDI value of 0.954 (Conceicao et al. 2019), categorizing it as "very high human development." It has a life expectancy of 82.3 years, a gender development index of 0.990, a gender inequality index of 0.044, is ranked 5th on the world's stage, and has a maternal mortality ratio of 5 per 100,000 deaths. In comparison, the United States has a gender inequality index of 0.182, is ranked 42nd, and has a maternal mortality ratio of 14 per 100,000 deaths. On the other hand, Niger has an HDI value of 0.377 (Conceicao et al. 2019), categorizing it as "low human development" and a life expectancy of 62 years. It has a gender development index of 0.298, a gender inequality index of 0.647, a maternal mortality ratio of 553 per 100,000 deaths, and is ranked 154th compared to other countries. Niger is a developing country in West Africa, and Norway a North Atlantic European country. This exploratory research aims to examine legacy population data from developed and underdeveloped global countries and examine possible reasons that impact the quality of life in these countries.

POPULATION GROWTH

Many factors influence population growth. According to Tejvan (2019), population growth is driven by fertility rates, viz, children per adult, and fatality rates. Both birth rates and mortality rates are, in turn, influenced by other factors such as economic growth and economic development. In many cases, these two factors have led to a decline in population growth. Other factors, such as availability of family planning, social expectations, and government intervention, can play an important role.

According to Minh (2012), Rapid population growth has outstripped increases in food production, and population pressure has led to the overuse of arable land and its destruction. Rapid growth has also hampered economic development and caused massive unemployment.

On the other hand, rapid growth in developed countries has outstripped increases in food production, and population pressure has led to the overuse of arable land and its destruction. Rapid growth has also hampered economic development and caused massive unemployment. (Gupta, Bongaarts and Cleland, 2011; Sinding, 2009)

So, what is the link between Population, Poverty, and Sustainable Development? Gupta, Bongaarts, and Cleland (2011) observed that public fund policy, population growth, the availability of natural resources, property resources, and family planning programs are all intricately linked and influence population growth, poverty levels, and sustainable development in both developed and developing countries.

METHODOLOGY

Population data for each year, starting in 2000 for both Norway and Niger until the present, were collected and analyzed. The numbers are given in millions. Then, a model for Norway and Niger was created using geometric sequences. A discrete set of values is separated each year. Each term is found by multiplying the previous term by a constant value. This standard constant value is defined as the average growth rate (also known as the ratio between the populations of consecutive years) signified by r and *R* in this case for both Norway and Niger.

The "average growth" is the average ratio between the populations of two consecutive years. In this case, it was determined for both Norway and Niger. The latter year (y_2) was divided by that of the previous year (y_1). For example, the year 2001 (11.751) was divided by 2000 (11.332) for Niger and received 1.037 as the growth rate. This results in a general formula: $y_2 \div y_1$. Then, adding the resulting growth rates and divided them by the number of terms (20) to receive my average growth rate for Niger and Norway of 1.039 and 1.009, respectively. The results (in addition to the calculated data and data from the resource) are shown below in Table 1 – Summary of Processed Data.

Edwards, Edwards, et al Population Growth

Year	# of years since 2000	Calculated Ratio	Norway Calculated Population (given in millions)	Norway Actual Population (given in millions)	Calculated Ratio	Niger Calculated Population (given in millions)	Niger Actual population (given in millions)
2000	0	-	4.499	4.499	-	11.332	11.332
2001	1	1.005	4.541	4.523	1.037	11.770	11.751
2002	2	1.005	4.584	4.546	1.037	12.226	12.190
2003	3	1.005	4.627	4.570	1.038	12.698	12.650
2004	4	1.006	4.670	4.598	1.038	13.190	13.126
2005	5	1.007	4.714	4.632	1.038	13.700	13.624
2006	6	1.009	4.758	4.673	1.038	14.230	14.144
2007	7	1.010	4.802	4.719	1.038	14.780	14.685
2008	8	1.011	4.847	4.771	1.039	15.352	15.251
2009	9	1.012	4.893	4.827	1.039	15.945	15.843
2010	10	1.012	4.939	4.886	1.039	16.562	16.464
2011	11	1.013	4.985	4.948	1.039	17.203	17.114
2012	12	1.013	5.032	5.014	1.040	17.868	17.95
2013	13	1.013	5.079	5.079	1.040	18.559	18.504
2014	14	1.012	5.126	5.142	1.040	19.277	19.240
2015	15	1.011	5.174	5.200	1.040	20.023	20.001
2016	16	1.010	5.223	5.251	1.039	20.797	20.789
2017	17	1.009	5.272	5.296	1.039	21.601	21.602
2018	18	1.008	5.321	5.338	1.039	22.437	22.442
2019	19	1.008	5.371	5.380	1.039	23.305	23.310
2020	20	1.008	5.421	5.421	1.038	24.206	24.206
The average growth rate for Norway		1.009					
The average growth rate for Niger		1.039					

Table 1 Summary of Processed Data

CALCULATIONS, ANALYSIS, AND INTERPRETATION:

This section presents the calculations, the analysis of the results, and our interpretations of the findings**.** This is appropriate because it uses a discrete set of values separated by years. Each term is found by multiplying the previous term by a constant value. This common constant value is defined as the average growth rate (also known as the ratio between the populations of consecutive years) signified by "r" and "*R,"* in this case for both Norway and Niger, respectively. This regression model is useful for this investigation because it explains the relationship between population growth and years.

Let (R) represent the average growth rate for Niger
Let (r) represent the average growth rate for Norway
Let (G) represent the Population (2000) of Niger in millions
Let (N) represent the Population (2000) of Norway in millions
Let (t) represent the time and number of years after 2000

Norway:

$$N_{1}\,(r)^{t} = N_{t}$$
$$4.499\,(1.009)^{t} = N_{t}$$

Niger

$$\boldsymbol{G_{1}\,(R)^{t} = G_{t}}$$
$$\boldsymbol{1.133\,(1.039)^{t} = G_{t}}$$

The difference in average growth ratio is not surprising because the Population of Niger grows significantly faster than that of Norway. This makes sense for the following reasons: Niger has a value that suggests that people are expected to attend school for a maximum of 6.5 years, but in actuality, on average, people go to school for 2.0 years. This means that the average number of years of Education received by Nigerien people ages 25 and older is 2 years. In addition, Niger's poverty rate remains 41.4%, affecting more than 9.5 million residents *(Overview, World Bank)*. This, combined with the fact that Niger's leading economy depends on agriculture, there will be more of a need for extra human labor. Therefore, more families will be inclined to produce more children, encouraging them to work in the fields and contribute to the Country's agricultural economy.

We plotted the data on a scatter graph to analyze the individual Country's population estimates and real-life data presented in figure 1. Combining the data on a single graph (figure 2), we analyzed and examined the results of both countries simultaneously and deciphered real-life situations.

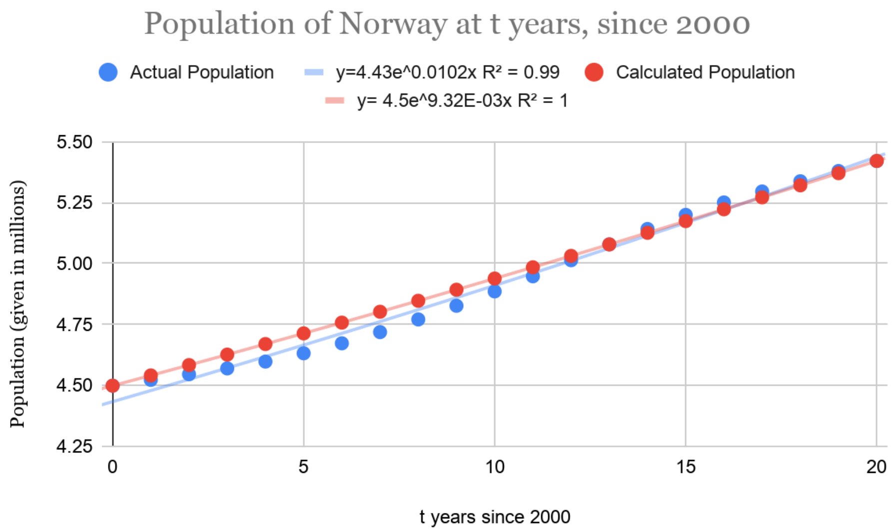


Figure 2 - Population of Norway at "t" Years Since 2000

Figure 2 shows the Calculated Population and the Actual Population of Norway in red and blue. At first glance, we can see that the Calculated Population has a continuous constant rate of change, whereas the Actual Population does not. This makes sense because a country's population does not increase consistently. Instead of forming a constant, undeviating function, the Actual Population appears to form an "S" curve for most of the 20-year timespan between 2001-2012 and 2014-2018.

This is further explained by the R^2 value, which reveals the strength of the correlation between the dependent and independent variables and how accurate the trend line equation ($y = 4.43e$ 〖^0.012x〗) models the data; the closer to 1, the stronger the correlation. The variables for this exploration are the independent (years since 2000) and dependent (population). With that information, while the R^2 value for the Actual Population equation is 0.99 and close to 1, it is not equal to 1. Therefore, the discrepancies shown in the years between 2001- 2012 and 2014 - 2018 are accurate and considered.

Multiple factors contribute to inconsistencies in the data, such as political unrest, war, famine, and terrorism, resulting in changes to the policy that may constrict a country's population growth. In this case, in the 2005/2006 timeframe, a decline in immigration and fewer births caused this inconsistency.

Norway has a constant, seemingly slow increase in population. This may be due to the gender equality present in Norway in contrast to Niger. Norway has a strong culture of feminism with a high gender development index of 0.990 and a low gender inequality index of 0.044. For comparison, Niger has a gender development index of 0.298 and a gender inequality index of 0.647. As a result, Norway prioritizes more gender equality than Niger. This leads to more women entering the workforce — as 77 percent of women make up Norway's workforce — or receiving an education. Cohen et al. (2011) showed that women who have advanced degrees have lower completed fertility in Norway. Either way, women in Norway have fewer children, approximately producing close to two children per year, and effectively maintaining the low Population of Norway.

The problem with examining population growth as a discrete geometric sequence is that population growth is continuous, not discrete. This is why we graphed the trend lines as exponential functions. The equation for the Calculated Population is $y = 4.5e$ 〖^(9.32E − 03x)〗

with an R^2 value of 1. From this information, we know that there is a much stronger correlation between the independent (years) and dependent variables (population) for the Calculated Population than the Actual Population.

This data for both Norway and Niger, as shown in figures 2 & 3, are being examined as continuous exponential functions.

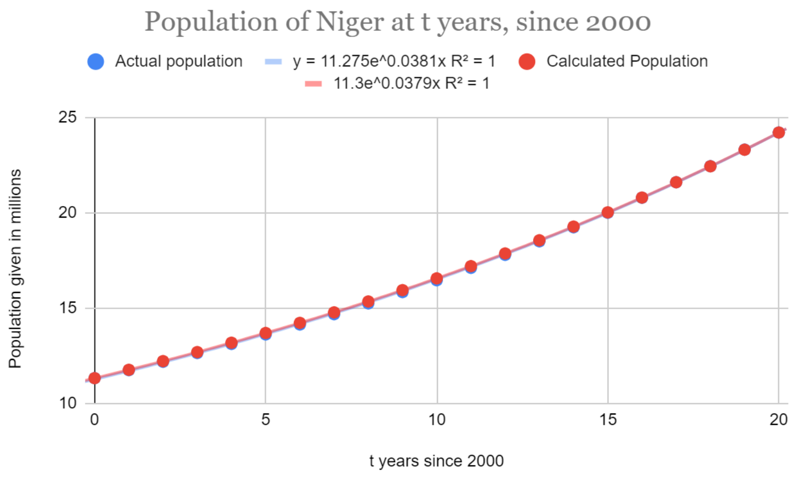


Figure 3 - Population of Niger at "t" Years Since 2000

We repeated the same process with the populations of Niger. The results were graphed to reflect the Actual Population and Calculated Population on a scatter chart graph (Figure 3) to examine their similarities and differences. Unlike Norway, Niger's Actual Population aligns closely with the calculated average growth rate (from Table 1 – Summary of Processed Data) and follows a continuous rate of change (Figure 3). In addition, the Actual Population (seen in blue) of Niger follows very closely to the calculated average percent growth rate, and thus the Calculated Population in red. Furthermore, the equations and R^2 values for both the trend lines and the R^2 values for Actual Population and the Calculated Populations are nearly identical: $y = 11.275e^{0.0381x}$ R^2= 1; $y = 11.3e^{0.0379x}$ R^2 = 1. The strong correlation with the data for Niger shows that there must have been minor population inconsistencies and changes in these 20 years, if any. This might be attributed to the constant need for more labor in the fields to fuel the agriculture economy. This is further revealed in Niger's fertility rate of "... close to 7 children per woman in 2016." (CIA World-Factbook). In fact, the total fertility rate is lower than the desired fertility rate. This means that the use of contraceptives is unlikely. It is this high fertility rate that sustains and allows for the Actual Population to meet the expected value of the Calculated Population and not deviate from the Calculated Population. Other factors such as gender inequality (with a gender inequality value of 0.647), lack of educational opportunities for women, absence of gynecological health care services, early marriage, and childbirth are also expected to contribute to women producing more children. The graph below (Figure 4) reflects both Country's data for Calculated Populations.

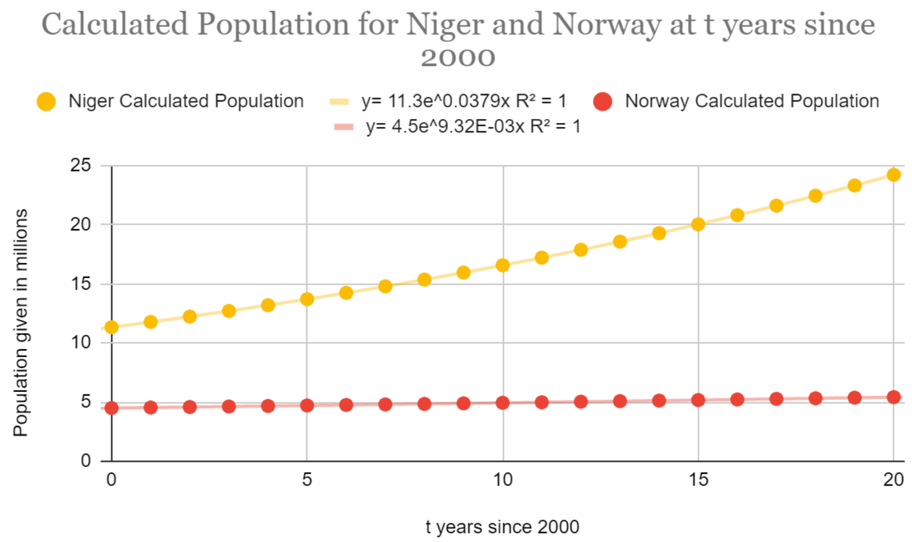


Figure 4 – Calculated Population for Niger and Norway at "t" Years Since 2000

Figure 4 shows the calculated population growth for both countries. There were two surprising characteristics of this graph: 1) the Population of Niger has a larger average percent growth rate (1.039) than Norway, and 2) the Population of Niger is significantly larger than the Population of Norway. Combining both data sets into one graph made it easier to compare both trend line equations modeling both Country's populations. For example, the Calculated Population for Niger has a trend line equation of $y = 11.3e \llbracket \hat{}0.0379x \rrbracket$ and an R^2 value of 1. In contrast, the Calculated Population for Norway has a trend line equation of $y = 4.5e \llbracket \hat{}(9.32E - 03x) \rrbracket$and an R^2 value of 1. While the R^2 values are exact, this only reveals the accuracy and the strength of the relationship between 2000 and the population growth. It is seen that as the years since 2000 increase, the population grows for both countries, each at their rate.

This may also be due to the level of urbanization for both countries. Norway's population is largely urban (83 percent) with a small percentage of arable land (only three percent); whereas, Niger has nearly 13 percent of arable and cultivated land and 16.6 percent of an urban population. Urbanization seems to be a factor that stifles population growth as it impacts the level of food production, as seen while examining Norway's Population.

According to Agarwal (2020), reflecting on English economist Thomas Malthus (1798), population growth is a geometric progression (an increase in which each term is found by multiplying the previous term by a constant number - a rate or ratio). However, factors such as food production are examined as an arithmetic progression (a progression or increase in a sequence because of a constant difference). Therefore, a country's population will grow faster, surpassing food production, diminishing as the Country reaches a specific capacity. This concept of population growth progression is especially true with Niger as "the dependence of most Nigeriens on subsistence farming on increasingly small landholdings, coupled with declining rainfall and the resultant shrinkage of arable land, are all preventing food production from keeping up with population growth." (CIA World-Fact-Book).

Ultimately, this will lead to starvation and the increase in the possibility that the population might diminish. Malthus' concept of population growth sheds light on other factors that impact population, such as land space. This led us to the question of "What is that capacity?" which may be a question that exceeds the limits of a geometric sequence. Therefore, with these questions in mind, two more areas of geometric calculations are used to determine the "capacity."

First, an examination of the projected Population of Norway and Niger in the year 2050, compared to that of a resource with the actual population statistics, revealed that using this simple formula to describe population growth suggested a more precise and accurate formula used. Presented below are the projected population for the year 2050 for Norway and Niger, using a geometric sequence.

$Norway$
$N_{1}(r)^{t} = N_{t}$
$2050 - 2000 = 50$
$4.499\,(1.009377342)^{50} = N_{50}$
$N_{50} = 7.175\ million$

Niger

$$\boldsymbol{G_{1}\,(R)^{t} = G_{t}}$$

$\mathbf{2050 - 2000 = 50}$
$\mathbf{11.332\,(1.039)^{50} = G_{50}}$
$\boldsymbol{G_{50} = 76.751\ million}$

Please also note that the numbers were converted to millions, i.e., 7.795 billion converted to 7,795 million.

$$N_{1}(r)^{(t-1)} = N_{t}$$
$$4.499\,(1.009377342)^{(t-1)} = 7{,}795$$
$(1.009377342)^{(t-1)} = 1732.607$
$$\log_{(1.009377342)} 1732.607 = t - 1$$
$$798.978 = t - 1$$
$$799.978 = t$$
$$t \approx 800\ years\ to\ reach\ world's\ population$$

$\boldsymbol{G_{1}\,(R)^{(t-1)} = G_{t}}$
$1.133\,(1.039)^{(t-1)} = 7.795$
$(1.039)^{(t-1)} = 6880.847$
$\log_{(1.039)} 6880.847 = t - 1$
$230.967 = t - 1$
$231.967 = t$
$It\ will\ take\ 232\ years\ to\ reach\ world's\ population\ \approx t$

These numbers are absurd and near impossible. For these countries to reach the world's population, these two countries will have to grow a landmass equivalent to all of the other countries in the world combined and provide sufficient resources such as food to sustain the population. However, as Agarwal (2020) Malthusian Theory Of Population - 1798, a country's population will surpass food production.

Finally, understanding the limitations of the geometric sequence and exponential function and limitations that impact population growth might be more prudent to examine the data through a more precise equation: the logistic equation ($y = K/(1 + Ae〚^(-kt)〛)$.) The logistic equation considers the limitations identified in the geometric sequencing calculations, such as the eventual plateau of the Population of Norway and Niger, due to limited access to natural resources (land). We did not use this formula to calculate the carrying capacity. This is an area for further exploration (a new study) that might use the logistic equation to model a new set of population data for Norway and Niger as a logistic equation.

The logistic equation is examined regarding Niger due to its low HDI value of 0.377 (Conceicao et al. 2019) and low human development category.

$$y = K/(1 + Ae〚^(-kt)〛)$$

Let (y) represent the calculated population
Let (K) represent the carrying capacity
Let (A) represent the initial Population of Niger
Let (a) represent the initial Population of Norway
Let (k) represents the rate of growth
Let (t) represent time

First, to find the carrying capacity of Niger, we must perform several steps:

1) Find the total land area of Niger, which is 1,266,700 sq km (CIA, World-Fact-Book).
2) Calculate the amount of land that will be used for inhabitation (building homes and residence). Since Niger has an agricultural economy, 13% of its land is arable. Thus, I will calculate the precise amount of arable land and subtract it from the total amount seen below.

$$13\%\ of\ 1{,}266{,}700\ sq\ km\ =\ x$$
$$0.13\ x\ 1{,}266{,}700\ = 164{,}671\ sq\ km\ of\ arable\ land$$
$$1{,}266{,}700 -\ 164{,}671 =\ 1{,}102{,}029\ sq\ km\ of\ inhabitable\ land$$

3) Assume that each family would need approximately 0.000223 sq km of land to live comfortably (Jpalmer, 2021), especially for large families, and divide the total amount of inhabitable land by this number, as shown below. This will give me us carrying capacity for Niger.

$$1{,}102{,}029\ \div\ 0.000223 = 494{,}183{,}4080$$
$$K \approx 490{,}000{,}000$$

CONCLUSION AND RECOMMENDATION

This research revealed that the population growth rate impacts countries differently. For developing countries, that impact is driven by Education, economics, and availability of arable land (Cohen, Joel E. E, et al, 2011). On the other hand, while some of those same factors are present for developed countries, there is less severity (Cohen, Joel E. E, et al, 2011).

Conclusion

Based on the findings of this research, it is estimated that by 2050, the Population of Norway is projected to be 6.600 million people (Norway Population), as opposed to the calculated 7.175 million people. Similarly, in 2050, the Population of Niger is projected to be about 65.593 million albeit, Niger's estimated population may not consider the influx of refugees entering Niger due to conflicts between Nigeria and Mali 2019 (Overview, World Bank).

Furthermore, the possibility of the number of COVID-19 deaths impacting subsequent population data was considered. However, it may be concluded that because Norway and Niger do not have an alarming number of COVID-19 deaths (649 and 185, respectively), the population is unlikely to diminish or plateau at this time significantly. By comparison, the United States has over 586 000 COVID-19 deaths, which will more likely considerably impact the United States Population.

The calculated numbers for Norway and Niger are unlikely to occur because an exponential model would only work indefinitely if there were an infinite number of resources. At a certain point, resources will begin to deplete (Boundless). Therefore, it is necessary to use a more accurate formula for population growth.

In conclusion, the data shows that it is important to examine population growth related to public health. Examining population growth reveals that the differences in cultural beliefs and values, living conditions, food production, and economic factors all play a significant role in determining the quality of life. It is indicative of why specific communities are better off than others due to access to resources. Although, the question remains: while 0.000223 sq km of land is projected to be sufficient for comfortable living, is this truly enough? What constitutes a good quality of life? This may serve as an area for further research.

Recommendations

The data did not address the fact women in Niger are producing more children due to the constant need for labor, and as a result, are not receiving adequate health services specific to women. Women centers for gynecological and reproductive health services, particularly in countries like Niger, must be established. These are countries with similar demographics and economies that depend on the fertility rate. In addition, health education in countries identical to Niger must also be mandated and encouraged to increase Education, promoting healthy living, particularly among women.

This research did not explore how the "carrying capacity" of Niger and Norway will better understand how resource limitations impact the quality of life in countries similar to Niger and Norway. Therefore, it is recommended that a comparative analysis of "carrying capacity" for both countries should be undertaken.

This harkens back to the original argument that motivated this study – how and what resources need to be redistributed to uplift underserved and marginalized communities in the U.S. in the following areas: Healthcare and Education. Healthcare and Education are fundamentally interwoven and contribute to health and well-being across all communities in the U.S. This is reflected in the COVID-19 deaths infection cases (Thebault, Tran, and Williams, 2020). In addition, according to Bhutta, Chang, Dettling, and Hsu (2020), "In the 2019 survey, White families have the highest level of both median and mean family wealth: $188,200 and $983,400, respectively. Black and Hispanic families have considerably less wealth than White families. Black families' median and mean wealth is less than 15 percent of White families, at $24,100 and $142,500, respectively. Hispanic families' median and mean wealth is $36,100 and $165,500, respectively". This explains the absence of educational resources in these communities.

Finally, it is recommended that analysis should be undertaken to explore the population growth further as it related to land space, specifically in regard to the number of years the populations of each country will take to "fill up" the Earth (in other words, "reach the world's population"). The entire world's population as of 2020 is 7.795 billion people, and it is only growing. Examining the number of years, the Population of Norway and Niger will reach the world's total population will approximate how people occupy space. In this case, "space" is the area that people on Earth currently inhabit. Furthermore, it will be necessary to examine the square meters of land inhabited by the people globally, but simply looking at the world's total population.

REFERENCES

"04227.Pdf." Montgomeryschoolmd.org. www.montgomeryschoolsmd.org/departments/regulatoryaccountability/glance/currentyear/schools/04427.pdf

Expansion of the Cropland Area:" Land Resource: Food Production, web.ccsu.edu/faculty/kyem/GEOG473/5thWeek/Food production, and GM foods.htm#:~:text=Malthusian Theory:&text=Malthus stated that population increased, it in a short time.

Agarwal, P. "Malthusian Theory Of Population." Intelligent Economist, 22 Nov. 2020, www.intelligenteconomist.com/malthusian-theory/#1-population-and-food-supply.

Bhutta, Chang, Dettling and Hsu "Disparities in Wealth by Race and Ethnicity in the 2019 Survey of Consumer Finances". Board of Govenors of the Federal Reserve System, 28 September 2020

Boundless. "Boundless Biology." *Lumen*, courses.lumenlearning.com/boundless-biology/chapter/environmental-limits-to-population-growth/#:~:text=Exponential growth may occur in,will plateau or level off.

Central Intelligence Agency, Central Intelligence Agency, www.cia.gov/the-world-factbook/countries/niger.

Cohen, Joel E. E, et al. Childbearing Impeded Education More than Education Impeded CHILDBEARING among Norwegian Women. 19 July 2011, www.pnas.org/content/108/29/11830.

Conceição, Pedro et al. “hdr2019.Pdf.” United Nations Development Programme, 2019. http://hdr.undp.org/sites/default/files/hdr2019.pdf

“Current World Population." *Worldometer*, www.worldometers.info/world-population/.

Developed Countries List 2021, worldpopulationreview.com/country-rankings/developed-countries.

“Finance & Development.” Finance & Development | F&D, www.imf.org/external/pubs/ft/fandd/basics/gdp.htm.

Gupta, Bongaarts and Cleland, "Population, Poverty, and Sustainable Development A Review of the Evidence", The World Bank Development Research Group Human Development and Public Services Team, 2011.

FocusEconomics. "What Is GDP?" *FocusEconomics | Economic Forecasts from the World's Leading Economists*, 29 Mar. 2014, www.focus-economics.com/economic-indicator/gdp#:~:text=GDP, short for Gross Domestic, Country in a given period.

"Jordvern." Regjeringen.no, www.regjeringen.no/no/tema/mat-fiske-og-landbruk/landbrukseiendommer/innsikt/jordvern/jordvern/id2009556/.

Jpalmer. "How Much Square Footage Fits Your Family?" Southdown Homes, 16 Feb. 2021, southdownhomes.com/how-much-square-footage-fits-your-family/#:~:text=The average living space required,range between $147,000 to $436,000.

Minh, Q.D. "Population and Economic Growth In Developing Countries", International Journal of Academic Research in Business and Social Sciences January 2012, Vol. 2, No. 1 (2012)

"Niger Population (LIVE)." Worldometer, www.worldometers.info/world-population/niger-population/.

"Niger: WHO Coronavirus Disease (COVID-19) Dashboard." World Health Organization, World Health Organization, covid19.who.int/region/afro/country/ne.

"Norway Population (LIVE)." *Worldometer*, www.worldometers.info/world-population/norway-population/.

"Overview." *World Bank*, www.worldbank.org/en/country/niger/overview#:~:text=Despite significant strides made by,more than 9.5 million people.

Ovind, Anne Kari H. "When A Government Makes Gender Equality A Top Priority, It Pays Off." HuffPost Canada, HuffPost Canada, 8 Mar. 2018, www.huffingtonpost.ca/anne-kari-h-ovind/norway-canada-gender-equality_a_23377768/

Roser, Max, et al. "Norway: Coronavirus Pandemic Country Profile." Our World in Data, 26 May 2020, ourworldindata.org/coronavirus/country/norway.

Sinding, S.W. "Population, poverty and economic development." Phil. Trans. R. Soc. B (2009) 364, 3023–3030 doi:10.1098/rstb.2009.0145

Tejvan Pettinger (Factors that affect population size and growth); https://www.economicshelp.org/blog/469/development/factors-effect-population-size-and-growth/ (2019)

Thebault, Tran and Williams. "The Coronavirus is Infecting and Killing Black Americans at an Alarmingly High Rate, 7 April, 2020

"UN Recognition of the Least Developed Countries." *UNCTAD*, unctad.org/topic/least-developed-countries/recognition.

"Why Do the Poor Have Large Families?" Compassion Australia, www.compassion.com.au/blog/why-do-the-poor-have-large-families.

DECISION SCIENCES INSTITUTE

Reconstructing Authentic Leadership: Uniting the Critical and Positive Divide

Catherine A. Helmuth
Central Michigan University
Email: helmu1c@cmich.edu

Sebastien Vendette
Central Michigan University
Email: vende1s@cmich.edu

Melissa L. Intindola
Bucknell University
Email: mli004@bucknell.edu

Chellie Spiller
University of Waikato Management School
Email: chellie.spiller@waikato.ac.nz

ABSTRACT

While academics and practitioners alike have celebrated authentic leadership's contributions, the literature is currently split between two perspectives – positive psychology versus critical scholarship. Our paper reviews positive psychology's dominant conceptualization and synthesizes a growing body of critical scholarship, which suggests authentic leadership needs a reconstruction. Rather than propagating divisional ideologies, we propose that authentic leadership's reconstruction should unite around the *process of becoming* as it coalesces key tenets from critical scholarship (existentialism, Jungian, narrative theory) and positive psychology. We conclude with the theoretical and practical benefits of authentic leadership's reconstruction.

KEYWORDS: Leadership, Authenticity, Authentic leadership, Critical Scholarship

DECISION SCIENCES INSTITUTE
Product Lifecycle, Manufacturing Strategy, and Innovation

Canchu Lin
SUNY Farmingdale State College
Email: linc@farmingdale.edu

ABSTRACT

This study addresses how manufacturing strategy shapes innovation at different stages of product lifecycle. Drawing on the strategic choice perspective and the product lifecycle literature, this study presents propositions on how each stage of product lifecycle determines what objective is prioritized over others in manufacturing strategy and how the prioritized objective shapes the focus of innovation. The paper then discusses how these propositions contribute to theory and practice.

KEYWORDS: Product lifecycle, manufacturing strategy, innovation, strategic choice, and contingency theory

INTRODUCTION

Prior research (e.g., Azadegan & Dooley, 2010) examined the effect of innovation on cost, quality, flexibility, and delivery from the organizational learning perspective. However, these four areas were treated as indicators of manufacturing performance. Yet, these four areas are widely accepted as constituents of manufacturing strategy, which is normally known as manufacturing priorities or objectives. According to the strategic choice perspective, emphasis on any of these four priorities in manufacturing strategy implies adoption or implementation of an innovation that aligns with the strategy choice (Chereau, 2015). Put it in another way, strategy choice would predict implementation of a certain type of innovation. Yet, there was rarely any attempt to explore this issue in past research. Further, strategy choice has been so far linked to business strategy in the literature. In the case of manufacturing strategy, choice usually means prioritizing one objective over others. It remains unexplored whether this type of prioritizing is stable or dynamic.

The purpose of this study is to fill these two gaps in operations management research. More specifically, we will draw on the contingency theory to argue that manufacturing strategy choice is a dynamic process, where prioritizing one objective over others changes from phase to phase in product lifecycle. On the basis of that, we then articulate another point that both prioritizing one objective over others and the phase in product lifecycle shape the choice of an innovation portfolio. Our overall argument is that innovation choice is aligned with product lifecycle and manufacturing strategy. We will frame this overall argument as a theoretical perspective and translate it into specific propositions. Doing that, we seek to contribute to the literature. The rest of the paper is outlined to accomplish these objectives. First, we will review the literatures of strategic choice, manufacturing strategy, product lifecycle, and innovation to build a theoretical foundation for the research model. Next, we will further draw on this theoretical foundation and specific literature to formulate the propositions. Lastly, we will conclude the paper with a short discussion of how this research will contribute to theory and practice.

LITERATURE REVIEW

Innovation as an Outcome of Strategic Choice in the Context of Product Lifecycle

One way to examine the relationship between strategy and innovation is the strategic choice approach (Chereau, 2015). Based on this approach, innovation behavior, activities, and then performance can be viewed as the outcome of strategic choice. Put it in another way, strategy predicts innovation behavior and activities (Kotabe, 1990; Zahra & Covin, 1994). This strategic choice approach also implies that only innovation attributes that are consistent with a firm's strategy can be expected to contribute to the firm's performance (Raymond & Saint-Pierre, 2010). Based on this strategic choice approach, in our investigation of the relationship between manufacturing strategy and innovation, it follows that innovation attributes are determined by choice of a manufacturing strategy. As manufacturing strategy is specifically reflected in the competitive priorities, we can reasonably argue that when one such priority is chosen over others, there must be certain innovation attributes that fit this priority.

While the strategic choice approach illuminates how innovation attributes are outcomes of choice of manufacturing competitive priorities, the contingency theory is a direction for how to explain strategic choice. Based on the contingency theory, contingent or situational variables shape strategic choice. In strategy research, prior studies strongly suggest that the product lifecycle (PLC) is a primary contingent variable determining strategic choice that leads to business performance (Anderson & Zeithaml, 1984, Hambrick & Lei, 1985; Hoffer, 1975). More specifically, Hoffer (1975) further viewed the stage of PLC is a market contingency variable. Although the strategic activities in each stage of the PLC proposed in these prior studies fall under the generic category of business strategy or a more specific category of marketing strategy, these studies provide important implications for identifying manufacturing strategic activities as suggested by the competitive priorities corresponding to the different stages of the PLC. In the following paragraphs, we review the concepts of PLC, Manufacturing strategy and competitive priorities, and innovation.

Product Lifecycle

Borrowed from biological sciences where living organisms are characterized, the term lifecycle is used in the business, especially marketing practice, to describe the evolution of product-markets. Like living organisms, a product has such a lifecycle, known as product lifecycle (PLC). The PLC is a descriptive framework that classifies the lifecycle of a product into distinguishable stages: introduction, growth, maturity, and decline (Levitt, 1965). During the *introduction* stage, the product is commercialized and launched to the market, where there are few competitors. As the product is more and more accepted in the market, it moves to the subsequent *growth* stage, in which sales revenues are growing rapidly and support costs decline as the product starts to become self-marketing. The increased sales also attract many new entrants into the market, creating strong competition. Advertising is used to promote the product. As sales growth begins to taper off, the product enters the *maturity* stage. In this stage, the product is at a point where its existence is never given a second thought. This is also the stage when market becomes more and more saturated, creating intense competition among firms for market share. When overall market sales begin to fall, the product enters the *decline* stage, when the product reaches the end of its life and is eventually withdrawn from the market. During this final stage, firms cut their marketing expenditures to reduce costs.

The utility of the PLC framework in guiding strategic planning has been well documented. For example, given its normative power, the PLC enables both practitioners and researchers to decide what strategies to implement under the different PLC stages (Thorelli & Burnett, 1981). More specifically, Hayes and Wheelwright (1979) emphasized that the lifecycle position of a firm's product strongly suggests what changes should be made in the firm's policies and procedures in their manufacturing function. More importantly, it guides the firm to accumulate customer requirements in the corresponding market condition of each stage and translate them into manufacturing needs (Hays & Wheelwright, 1984).

Because of its perceived role in strategic planning, prior research examined how PLC relates to manufacturing strategy, more specifically, the competitive priorities (see, e.g., Magnan, Fawcett, & Birou, 1999). Similarly, the literature also supplies us with enough empirical studies investigating how PLC relates to innovation, e.g., shorter PLCs found in parallel to higher intensity in product and process innovation in certain industries such as seed corn (Magnier, Kalaitzantonakes, & Miller, 2010), and shorter PLCs matched with shorter time to market, time to volume, and time to breakeven, three indicators of new product development innovation performance (Plewa, 2016). Yet few studies, if not at all, examined the dynamics of PLC, manufacturing strategy, and innovation in a single project. We argue that there is much merit of examining these three concepts together as we believe that PLC shapes manufacturing strategy choice, which then triggers appropriate innovation practice. Although weaknesses have been identified in the PLC framework, its heuristic power is helpful to scholarly examination of product-market dynamics (Wong & Ellis, 2007). We argue that this heuristic power can be extended to investigate manufacturing strategy in its entirety especially in the context of product dynamics. This particularly benefits our examination of the dynamics of the competitive priorities in that it helps to reveal the rise and fall of the priorities in strategic importance at the different stages of the product lifecycle, which then predict a firm's engagement in specific types of innovation that is critical to their accomplishment of their strategic priorities.

Manufacturing Strategy and Competitive Priorities

Only when enough strategic importance is attached to manufacturing, the impact of manufacturing on general organizational performance can be appropriately recognized (Skinner, 1969). For example, Acquaah, Amoako-Gyampah, and Jayaram (2011) found that a sound manufacturing strategy helps a firm to develop a defense as well as resilience against economic disruption. More specifically, implementing a manufacturing strategy that best matches manufacturing objectives helps to generate superior outcomes (Devaraj, Hollingworkth, & Schroeder, 2004). Further, best manufacturing practice can be attributed to having a sound manufacturing strategy (Voss, 2005). Moreover, the impact of manufacturing strategy even extends to the area of customer focus practices (Sousa, 2003). Similarly, exercising manufacturing strategy influences workforce development, besides directly effecting product-process development (Swink, Narasimahan, & Kim, 2005).

Just as a firm should have an overall competitive strategy (Porter, 1980), it should have a matching manufacturing strategy (Joshi, Kathuria, & Porth, 2003; Swink, Narasimhan, & Wang, 2007; Ward & Duray, 2000). Based on Chatha and Butt's (2015) recent review of the manufacturing strategy literature, an overwhelming majority of the studies reviewed treated the competitive priorities as the major component of manufacturing strategy, among its multiple definitions. At least, they are the content elements of manufacturing strategy (Swamidass & Newell, 1987). This is because these priorities or dimensions win orders for firms (Hill, 1994), and serve as a basis for firm strategic decision making and reflect its manufacturing mission and policies (Schroeder, Anderson, & Cleveland, 1986). Moreover, these competitive priorities serve as a linkage between competitive strategy and manufacturing performance (Amoako-Gyampah & Acquaah, 2008; Kim & Arnold, 1996; Ward & Duray, 2000). Even though there is an ongoing debate about whether a firm just focuses on one priority for achieving superior performance or can target on multiple priorities simultaneously (Schroeder, Shah, & Peng, 2011), these priorities are of great strategic importance to the firm (Skinner, 1969). Further, these priorities are even deemed as capabilities that firms must develop to enhance their competitive advantage (Ferdows & De Meyer, 1990). Thus, this is a mainstream view that the competitive priorities constitute the core part of manufacturing strategy. In empirical research, these priorities are used to operationalize manufacturing strategy (see, e.g., Youndt, Snell, Dean, & Lepak, 1996). These competitive priorities include cost, quality, delivery, and flexibility (Hayes & Wheelwright, 1984; Van Dierdonck & Miller, 1980; Wheelwright, 1984).

Cost as a competitive priority refers to a firm's emphasis on reducing production costs and inventory, increasing facilities and capacity utilization (Ward & Duray, 2000). The priority of *quality* is the important quality aspect of a firm's process control and process management (Flynn, Schroeder, & Sakakibara, 1994). More specifically, it means that to enhance its quality control practice, the firm emphasizes on implementing statistical process control methods and real time process control systems, updating process equipment, and developing new processes for new and old products (Ward & Duray, 2000). *Flexibility* is a firm's capability of making quick changes, based on customer requirement, in production volume and product mix. When treating flexibility as a competitive priority, a firm usually emphasizes on reducing lead time and setup time, changing priority of jobs and machine assignments on the job shop floor (Gerwin, 1993). Recently, more dimensions have been added to manufacturing flexibility. For example, Rogers, Ojha, and White (2011) established six dimensions of flexibility, which include product mix flexibility, routing flexibility, equipment flexibility, volume flexibility, labor flexibility, and supply management flexibility. The *delivery* priority refers to a firm's emphasis on customer service by enhancing its delivery reliability and speed (Ward & Duray, 2000).

A significant number of studies explored and affirmed the relationship between these priorities and firm performance. However, innovation, which also positively contributes to manufacturing performance (Adebanjo, The, & Ahmed, 2018), is, surprisingly, not included among these competitive priorities in the manufacturing strategy literature (Chatha & Butt, 2015). A sound explanation for this can be that innovation is a means through which a competitive priority contributes to firm performance. For this reason, we will survey the innovation literature in the following section.

Innovation

The innovation literature supplies multiple typologies of innovation (for example, incremental vs. radical, exploratory vs. exploitative, technology-based or market-based, etc.), which are developed based on diverse theoretical perspectives. From a strategy point of view, regardless of its strategic orientation, a firm's innovative performance is usually shown in product, process, marketing, and organizational innovation (Chereau, 2015). As this study aims to examine the relationship between manufacturing strategy and innovation in the context of product lifecycle, which is the market, these four types of innovation are used to denote firm's innovation performance.

The market dynamics of PLC suggests that customer demand is a driving force behind the market change and the associated changes in the product over time. Thus, for firms, their innovation must be shown in addressing customer demand in general, and more specifically, in learning about and capturing customer needs, developing new products to meet these needs and creating and seeking new processes to improve their products, and developing and implementing new organizational structure that facilitates all those changes (Chereau, 2015). Thus, the PLC perspective forecasts activities of marketing, product and process, and organizational innovation in firms.

Multiple definitions of these four types of innovation exist in the innovation literature. Among them, those from the Organization for Economic Co-operation and Development (OECD) are widely accepted (Hullova, Trott, & Simms, 2016). In their definition, *product innovation* refers to development and implementation of a good or service that is significantly improved in technical specifications, composition of elements, software attached or embedded in the product, product interface or other functional characteristics. Next, *process innovation* is defined as development or implementation of a production or delivery method that has significant improvements in techniques, equipment and/or software. Similarly, *marketing innovation* refers to development or adoption of a new marketing method that involves “significant changes in product design or packaging, product placement, product promotion or pricing.” Lastly, *organizational innovation* is defined as creating or implementing “a new organizational method in business practices, workplace organization or external relations.” (OECD, 2018).

Among these four types of innovation, product and process innovation have been inherent in the conceptual framework of product lifecycle (PLC) (Abernathy & Utterback, 1978; Klepper, 1996). Although empirical studies have documented the relationship between product and process innovation and PLC, most of these studies just showed a general trend in a certain industry. For example, the trend of shorter PLCs is related to that of accelerated levels of biotech product innovation in the US seed corn industry (Magnier, Kalaitzandonakes, & Miller, 2010). Few studies have attempted to examine how product and process innovation emerge in the different PLC phases, especially as a consequence of a manufacturing strategy choice. Similarly, a significant number of studies investigated the complementarity of product and process innovation (Hullova, Trott, & Don Simms, 2016), but this complementarity has been rarely examined in the context of PLC evolution.

Next, although the concept of PLC explicates the evolution of a product, the change in the product over time is related to market change. Marketing innovation is critical to a firm’s entry into and growth in new markets (Schubert, 2010; Varis & Littunen, 2010). Likewise, it enhances a firm’s competitiveness in a foreign market (Gupta, Malhotra, Czinkota, & Foroudi, 2016). Thus, marketing innovation, like product and process innovation, must be inherent in the conceptual framework of PLC. Further, marketing innovation relates to strategy (Menon, Bharadwaj, Adidam, & Edison, 1999; Verhoef & Leeflang, 2009), and more specifically to manufacturing strategy (Chereau, 2015). Yet, it is rarely examined in the context of PLC as well.

Finally, organizational innovation is akin to what is already known as administrative innovation based on their definitions. According to Damanpour (1991), administrative innovation is only indirectly related to work activities and mainly deals with administrative processes including those for doing business practice, organizing people, and reaching out to other stakeholders, etc., which organizational innovation is concerned with. Thus, in a way, organizational innovation, based on the OECD definition, is embedded in administrative innovation. At any phase of the PLC, firms institute certain organizational processes that facilitate their accomplishment of the chosen manufacturing strategy. For example, creating or adopting a new organizational structure that is characterized by high or low formalization and centralization would result in corresponding administrative processes (Daft, 1978), which may be conducive to the chosen manufacturing strategy. For this reason, organizational innovation relates to manufacturing strategy in the context of PLC.

PROPOSITIONS

Competitive Priorities and Innovation

Quality

Quality management as an organizational strategic focus serves as an important foundation for innovation and more importantly, triggers the innovation process (McAdam & Armstrong, 2001). Empirical research already supported this view. For example, both Flynn (1994) and Kim, Kumar, and Kumar (2012) showed that quality management is positively related to product and process innovation. Similarly, Hung, Lien, Yang, Wu, and Kuo (2011) found that total quality management (TQM) is strongly related to product and process innovation. Both product and process innovation centers around product development, but more specifically involves translating customer needs into attributes of the product (Hung et al., 2011). This must be most apparent in the *introduction* stage.

Both quality management and TQM refer to a management approach that reflects managerial emphasis of and guidance to quality management. In the manufacturing function, quality management and TQM are best shown in the competitive priority of quality. Zeng, Phan, and Matsui (2015) showed that the hard but not the soft aspect of quality management is positively related to product and process innovation. The hard aspect refers to the specific quality management tools, techniques, and processes, which are most likely used in the manufacturing function. This helps to bolster the view that the quality competitive priority is strongly associated with product and process innovation. As TQM is also strongly related to adopting new internal organizational management methods such as HR practices and new or flexible organizational structure (Hung et al., 2011), it triggers organizational innovation as well. Further, centralization of authority and functional integration, both being areas of organizational innovation, impact quality management's relationship to product innovation (Zeng, Zhang, Matsui, & Zhao, 2017). This further shows that the competitive priority of quality is associated with organizational innovation.

Flexibility

As the need of prioritizing flexibility in manufacturing stems from meeting customer needs, a key concern of marketing innovation, flexibility as a competitive priority and marketing innovation must be a good fit. In other words, the competitive priority of flexibility must strategically motivate marketing innovation. Marketing innovation can be shown in such activities as doing innovative market research to best capture and sometimes cultivate customer needs, and using new methods to communicate to customers that the product meets their needs.

Additionally, manufacturing flexibility achieves the goals of meeting customer needs and still contributing to performance mainly through product, process, and organizational innovation (Camison & Lopez, 2010). First, the goal of meeting customer needs pushes the firm to seek changes or innovations in product design so that the product maximally reflect and then actualize these needs. Doing that, the firm engages itself in product innovation, through such activities as innovative design, purchasing materials that enhance the functions of the product that reflect customer needs, and seeking supplier involvement (Tavani, Sharifi, & Ismail, 2014) in responding to changes in product mix. Empirical research shows that manufacturing flexibility has a positive association with product innovation (Menor, Kristal, & Rosenweig, 2007; Suarez, Cusumano, & Fine, 1996), which contributes to firm performance (Bayus, Erickson, & Jacobson, 2003; Gopalakrishnan & Damanpour, 1997).

Next, it is very obvious that manufacturing flexibility is enabled by process innovation, i.e., use and adoption of new production processes (Camison & Lopez, 2010), which are used, for example, to reduce lead time and setup time, and change priority of jobs and machine assignments on the job shop floor (Gerwin, 1993). Both versatile machinery and versatile workforce are needed in quickly responding to fluctuations in manufacturing and other processes (Duguay, Landry, & Pasin, 1997). Only when innovation is introduced, these processes will likely improve (Camison & Lopez, 2010). More specifically, to achieve volume flexibility, for example, innovation in processes of supplier involvement, facilities ulitization, and labor and machine scheduling is crucial.

Lastly, manufacturing flexibility also calls for organizational innovation. To enhance manufacturing flexibility, product and process innovation must be accompanied by organizational innovation. Product and process innovation loom in organizations that cultivate employee commitment and productivity with a right organizational structure. Organizations emphasizing flexibility as a competitive priority usually take on a decentralized structure and use teams to organize their work, where members tend to be multi-skilled and flexible (Duguay et al., 1997). Likewise, organizations that prioritize manufacturing flexibility tend to establish innovative methods in performing external relationship building activities such as seeking cooperation with customers, suppliers, or research centers (Camison & Lopez, 2010). Transforming an organization into a decentralized structure and using new methods to do external relationship building are examples of organizational innovation, based on the OECD's defition. Organizational innovation, just as product and process innovation, facilitates the relationship between manufacturing flexibility and firm performance (Camison & Lopez, 2010).

Delivery

Extant literature has rarely addressed how delivery relates to innovation. Among the four priorities, delivery follows flexibility in meeting customer needs as a major concern. While flexibility emphasizes on meeting customer needs in terms of their preferences for attributes of the product and requirements of different demands over time, delivery mainly refers to speed and dependability of transporting the product to customers. Hence, among the four types of innovation, process and organizational innovation are more likely to strengthen delivery. Delivery itself is an organizational process and hereby relates more to process and organizational innovation than to marketing and product innovation. Innovative delivery methods or process would directly lead to significant improvement in delivery speed and dependability. Similarly, adopting new and innovative technologies or other means to shorten the production process would also result in increase in delivery speed and dependability. Likewise, innovative organizational methods for improving workplace communication and organization would also lead to improvement in organizational processes including delivery.

Cost

Prior research has investigated whether business practices such as offshoring and outsourcing that are primarily concerned with reducing cost impact innovation (Fifarek, Veloso, & Davidson, 2008), but rarely directly addressed how cost as a competitive priority would trigger innovation in the firm. As production efficiency and economies of scale target on cost reduction, process innovation, among the four, is the most likely means for a firm to address the cost priority. Following process innovation, product innovation will also be sought by the firm to accomplish this objective, as it will lead to new designs or new materials used for the product which will help to reduce cost as an outcome. Next, as organizational innovation can play an auxiliary role in both product and process innovation, the firm can also rely on it as a means to reduce cost. Hence, organizational innovation can be an outcome of the cost priority for innovation.

Innovations as Outcomes of Competitive Priorities at the Stages of PLC

In the spirit of the contingency theory, product evolution over time predicts corresponding changes in the competitive priorities that govern manufacturing performance (Hayes & Wheelwright, 1984). Changes in the competitive priorities can be characterized as what priority rises to be the order-winner and what priorities serve as order-qualifiers (Hill, 1994). An order-winner is the priority that determines the most important element of the product from the customer's perspective, whereas an order-qualifier is one that directs the firm to improve certain attributes of the product that help to enhance its attractiveness to the customer.

Introduction

At the *introduction* stage, the manufacturer's major concern is product development (Hoffer, 1975). Thus, continuous improvement of the quality of the product is the leading competitive priority. Besides, as the *introduction* stage strategically emphasizes a buyer focus (Anderson & Zeithaml, 1984), which is to investigate and satisfy buyer needs beside the quality need and purchasing frequency (Hoffer, 1975), the firm should seek to quickly respond to these needs. This is where the priorities of flexibility and delivery are concerned about. Thus, at the introduction stage, the order winner is quality, and the order qualifiers are: cost, flexibility, and delivery. This distribution has received empirical support (Magnan et al., 1999). Thus, our first proposition can be established as:

H1: At the introduction stage, firms are mainly engaged in product innovation, process innovation, and organizational innovation to address their competitive priority of quality. Next, they will seek: process innovation, product innovation, and organizational innovation to address their competitive priority of cost; marketing innovation, product innovation, process innovation, and organizational innovation to address their competitive priority of flexibility; and process innovation, and organizational innovation to address their competitive priority of delivery.

Growth

The crucial concern of the *growth* stage is the product performance in relation to customer needs and thereby continuous improvement of the product so that it meets customer needs (Anderson & Zeithaml, 1984). By definition, meeting the changing needs of customers is the priority of flexibility. Thus, flexibility is the order winner of this stage. Next to flexibility in importance, delivery is also crucial to meeting customer needs. So, delivery follows flexibility as the first order qualifier for this stage. In addition, other strategic variables at this stage include product modification and improvement (Hay & Ginter, 1979; Wind, 1981), and economies of scale, production efficiency, and price reduction (Rumelt, 1979). These variables translate into the priorities of quality and cost. Thus, the competitive priorities at the growth stage take on the following distribution: flexibility as order winner, and delivery, quality, and cost as order qualifiers. This distribution is also empirically supported (Magnan et al., 1999). Based on these characteristics, we advance our proposition for the growth stage here:

H2: At the growth stage, firms are mainly engaged in marketing innovation, product innovation, process innovation, and organizational innovation to address their competitive priority of flexibility. Next, they will seek: process innovation and organizational innovation to address their competitive priority of delivery; product innovation, process innovation, and organizational innovation to address their competitive priority of quality; and process innovation, product innovation, and organizational innovation to address their competitive priority of cost.

At this stage, their marketing innovation should focus on seeking customer preference of their product instead of getting them to try it (Levitt, 1965).

Maturity

For the *maturity* stage, strategic recommendation focuses on cost reduction coupled with quality improvement (Hall, 1980), price reduction (Hay & Ginter, 1979), meeting buyer needs (Hoffer, 1975), product differentiation (Hay & Ginter, 1979), and emphasizing on product quality (Hamermesh & Silk, 1979). In summary, the major concerns of the maturity stage take the following order: process efficiency, reducing costs in marketing and distribution, further product and market differentiation in terms of customer needs, and increased quality (Anderson & Zeithaml, 1984). Thus, at the maturity stage, cost rises to be the order winner for firms, and delivery, flexibility, and quality serve as their order qualifiers. Such a distribution of the competitive priorities has also received empirical support (Magnan et al., 1999). Summarizing these findings, we propose our proposition for the maturity stage as:

H3: At the maturity stage, firms are mainly engaged in process innovation, product innovation, and organizational innovation to address their competitive priority of cost. Next, they will seek: process innovation and organizational innovation to address their competitive priority of delivery; marketing innovation, product innovation, process innovation, and organizational innovation to address their competitive priority of flexibility; and product innovation, process innovation, and organizational innovation to address their competitive priority of quality.

Decline

Finally, for the *decline* stage, Hoffer (1975) recommended the following strategies: buyer loyalty, degree of product differentiation, price elasticity of demand, and product quality. Both buyer loyalty and product differentiation point to flexibility as a competitive priority, whereas price elasticity of demand and product quality relate to the priorities of cost and quality respectively. Thus, for the decline stage, flexibility is the order winner for firms, and cost, quality, and delivery are their order qualifiers. Such a figuration of the competitive priorities is supported in empirical research (Magnan et al., 1999). These findings prompt us to propose our proposition for the decline stage as:

H4: At the decline stage, firms are mainly engaged in marketing innovation, product innovation, process innovation, and organizational innovation to address their competitive priority of flexibility. Next, they will seek: process innovation, product innovation, and organizational innovation to address their competitive priority of cost; product innovation, process innovation, and organizational innovation to address their competitive priority of quality; and process innovation and organizational innovation to address their competitive priority of delivery.

CONCLUSION

Prior research treated innovation as a determinant of manufacturing performance (Adebanjo, The, & Ahed, 2018; Azadegan & Dooley, 2010). However, innovation occurs in multiple organizational processes including production, marketing, and administration (Chereau, 2015). Thus, firms may focus on innovation in certain processes over others. We argue that innovation focus is an outcome of strategic choice and this is especially so at different stages of product lifecycle. With theoretical support from the strategic choice perspective and the research literature of product lifecycle, we then extend this overall argument to articulate specific propositions about how innovation should be aligned with manufacturing strategy at different stages of product lifecycle. These propositions collectively contribute to theoretical development in this area of research that otherwise remains a stalemate (Chatha & Butt, 2015). No less important, they will guide managerial practice with regard to how innovation should be aligned with manufacturing strategy.

REFERENCES

Abernathy, W. J., & Utterback, J. M. (1978). Patterns of industrial innovation. *Technology Review, 80*, 41-47.

Acquaah, M., Amoako-gyampah, K., & Jayaram, J. (2011). Resilience in family and nonfamily firms: An examination of the relationships between manufacturing strategy, competitive strategy and firm performance. *International Journal of Production Research,*

Adebanjo, D., Teh, P. L., & Ahmed, P. K. (2018). The impact of supply chain relationships and integration on innovative capabilities and manufacturing performance: The perspective of rapidly developing countries. *International Journal of Production Research, 56(4)*, 1708-1721.

Amoako-Gyampah, K., & Acquaah, M. (2008). Manufacturing strategy, competitive strategy and firm performance: An empirical study in a developing economy environment. *International Journal of Production Economics, 111*, 575–592.

Anderson, C. R., & Zeithaml, C. P. (1984). Stage of the product lifecycle, business strategy and business performance. *Academy of Management Journal, 27 (1)*, 5-24.

Azadegan, A., & Dooley, K. J. (2010). Supplier innovativeness, organizational learning styles and manufacturer performance: An empirical assessment. *Journal of Operations Management, 28*, 488-505.

Bayus, B. L., Erickson, G., & Jacobson, R. (2003). The financial rewards of new product introductions in the personal computer industry. *Management Science, 49 (2)*, 197-210.

Camison, C., & Lopez, A. V. (2010). An examination of the relationship between manufacturing flexibility and firm performance: The mediating role of innovation. *International Journal of Operations and Production Management, 30 (8)*, 853-878.

Chereau, P. (2015). Strategic management of innovation in manufacturing SMEs: Exploring the predictive validity of strategy-innovation relationship. *International Journal of Innovation Management, 19(1)*, 1-37.

Daft, R. L. (1978). A dual-core model of organizational innovation. *Academy of Management Journal, 21*, 193-210.

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of Management Journal, 34 (3)*, 555-590.

Devaraj, S., Hollingworth, D. G., & Schroeder, R. G. (2004). Generic manufacturing strategies and plant performance. *Journal of Operations Management, 22 (3)*, 313–333.

Duguay, C. R., Landry, S., & Pasin, F. (1997). From mass production to flexible/agile production. *International Journal of Operations & Production Management, 17 (2)*, 1183-1195.

Ferdows, K., & De Meyer, A. (1990). Lasting improvements in manufacturing performance: In search of a new theory. *Journal of Operations Management, 9 (2)*, 168–184.

Fifarek, B. J., Veloso, F. M., & Davidson, C. I. (2008). Offshoring technology innovation: A case study of rare-earth technology. *Journal of Operations Management, 26*, 222-238.

Flynn, B. B. (1994). The relationship between quality management practices, infrastructure and fast product innovation. *Benchmarking Quality Management Technology, 1 (1)*, 48-64.

Flynn, B. B., Schroeder, R., & Sakakibara, S. (1994). A framework for quality management and an associated measurement instrument. *Journal of Operations Management, 11*, 339–366.

Fox, H. W., & Rink, D. R. (1977). Coordination of purchasing with sales trends. *Journal of Purchasing and Materials Management, 13 (4)*, 10-18.

Gerwin, D. (1993). Manufacturing flexibility: A strategic perspective. *Management Science, 39*, 395–410.

Gopalakrishnan, S., & Damanpour, F. (1997). A review of innovation research in economics, sociology and technology management. *Omega, 25 (1)*, 15-28.

Gupta, S., Malhotra, N. K., Czinkota, M., & Foroudi, P. (2016). Marketing innovation: A consequence of competitiveness. *Journal of Business Research, 69*, 5671-5681.

Hall, W, K. (1980). Survival strategies in a hostile environment. *Harvard Business Review,58 (5)*, 75-85.

Hambrick, D. C., & Lei, D. (1985). Toward an empirical prioritization of contingency variables for business strategy. *Academy of Management Journal, 28 (4)*, 763-788.

Hay, R. D., & Ginter, P. M. (1979). Strategies for maintaining a share of the market. Paper presented at the Annual Meeting of the Southern Academy of Management, Atlanta.

Hayes, R. H., & Wheelwright, S. C. (1979). Link manufacturing process and product lifecycle. *Harvard Business Review (Jan-Feb)*, 133-140.

Hayes, R. H., & Wheelwright, S. C. (1984). *Restoring our competitive edge: Competing through manufacturing*. New York: John Wiley and Sons.

Hayes, R. H., Wheelwright, S. C., & Clark, K. (1988). *Dynamic manufacturing*. New York: The Free Press.

Hill, T. J. (1994). *Manufacturing strategy: Text and cases, 2nd ed*. Burr Ridge, IL: Irwin.

Hoffer, C. (1975). Toward a contingency theory of strategy. *Academy of Management Journal, 18*, 784-810.

Hullova, D., Trott, P., & Simms, C. D. (2016). Uncovering the reciprocal complementarity between product and process innovation. *Research Policy, 45*, 929-940.

Hung, R. Y. Y., Lien, B. Y. H., Yang, B., & Kuo, C. M. (2011). Impact of TQM and organizational learning on innovation performance in the high-tech industry. *International Business Review, 20*, 213-225.

Joshi, M. P., Kathuria, R., & Porth, S. J. (2003). Alignment of strategic priorities and performance: An integration of operations and strategic management perspectives. *Journal of Operations Management*, *21(3)*, 353-369.

Kim, D. Y, Kumar, V., & Kumar, U. (2012). Relationship between quality management practices and innovation. *Journal of Operations Management, 30 (4)*, 295-315.

Kim, J. S., & Arnold, P. (1996). Operationalizing manufacturing strategy: An exploratory study of constructs and linkage. *International Journal of Operations & Production Management, 16,* 45–73.

Klepper, S. (1996). Entry, exit, growth, and innovation over the product lifecycle. *The American Economic Review, 86 (3)*, 562-583.

Kotabe, M. (1990). Corporate product policy and innovation behavior of European and Japanese multinationals: An empirical investigation. *Journal of Marketing, 54 (2)*, 19-33.

Levitt, T. (1965). Exploit the product lifecycle. *Harvard Business Review, 18 (November-December)*, 81-92.

Magnan, G. M., Fawcett, S. E., & Birou, L. M. (1999). Benchmarking manufacturing practice using the product lifecycle. *Benchmarking: An International Journal, 6 (3)*, 239-253.

Magnier, A., Kalaitzantonakes, N., & Miller, D. J. (2010). Product lifecycles and innovation in the US seed corn industry. *International Food and Agribusiness Management Review, 13 (3)*, 17-36.

McAdam, R, & Armstrong, G. (2001). A symbiosis of quality and innovation in SMEs: A multiple case study analysis. *Managerial Auditing Journal, 16 (7)*, 394–399.

Menon, A., Bharadwaj, S. G., Adidam, P. T., & Edison, S. W. (1999). Antecedents and consequences of marketing strategy making: A model and a test. *Journal of Marketing, 63*, 18-40.

Menor, L., Kristal, M.M., & Rosenweig, E. (2007). Examining the influence of operational intellectual capital on capabilities and performance. *Manufacturing and Service Operations Management, 9 (4)*, 559-578.

OECD (2018): Better Policies for Better Lives. OECD Innovation Strategy: Defining Innovation. Retrieved from: http://www.oecd.org/site/innovationstrategy/defininginnovation.htm.

Plewa, M. (2016). Long-run dynamics between product lifecycle length and innovation performance in manufacturing. *International Journal of Innovation Management, 21 (1)*, 1750006: 1-20.

Porter, M. E. (1980). *Competitive strategy*. New York: The Free Press.

Raymond, L., & Saint-Pierre, J. (2010). Strategic capabilities for product innovation in SMEs: A gestalts perspective. *Entrepreneurship and Innovation, 11 (3)*, 209-220.

Rogers, P. P., Ojha, D., & White, R. E. (2011). Conceptualizing complementarities in manufacturing flexibility: A comprehensive view. *International Journal of Production Research 49 (12)*, 3767–3793.

Rumelt, R. P. (1979). Evaluation of strategy: Theory and models. In D. E. Schendel & C. W. Hofer (Eds.), *Strategic management: A new view of business policy and planning, 196-212.* Boston: Little, Brown.

Schroeder, R. G., Anderson, J. C., & Cleveland, G. (1986). The content of manufacturing strategy: An empirical study. *Journal of Operations Management, 6 (4)*, 405-415.

Schroeder, R. G., Shah, R., & Peng, D. X. (2011). The cumulative capability ‘sand cone’ model revisited: A new perspective for manufacturing strategy. *International Journal of Production Research, 49 (16)*, 4879-4901.

Schubert, T. (2010). Marketing and organizational innovations in entrepreneurial innovation processes and their relation to market structure and firm characteristics. *International Journal of Industrial Organizations, 36*, 189-212.

Skinner, W. (1969). Manufacturing — missing link in corporate strategy. *Harvard Business Review, 47(3)*, 136–145.

Sousa, R. (2003). Linking quality management to manufacturing strategy: An empirical investigation of customer focus practices. *Journal of Operations Management, 21 (1)*, 1–18.

Suarez, F. F., Cusumano, M. A., & Fine, C. H. (1996). An empirical study of manufacturing flexibility in printed circuit board assembly. *Operations Research, 4 (1)*, 223-240.

Swamidass, P. M., & Newell, W. T. (1987). Manufacturing strategy, environmental uncertainty and performance: A path analytical model. *Management Science, 33(4)*, 509-524.

Swink, M., Narasimahan, R., & Kim, S. W. (2005). Manufacturing practices and strategy integration: Effects on cost efficiency, flexibility, and market-based performance. *Decision Sciences, 36 (3)*, 427–457.

Swink, M., Narasimhan, R., & Wang, C. (2007). Managing beyond the factory walls: Effects of four types of strategic integration on manufacturing plant performance. *Journal of Operations Management, 25 (1)*, 148-164.

Tavani, S. N., Sharifi, H., & Ismail, H. S. (2014). A study of contingency relationships between supplier involvement, absorptive capability, and agile product innovation. *International Journal of Operations and Production Management, 34 (1)*, 65-92.

Thorelli, H. B., & Burnett, S. C. (1981). The nature of product lifecycles for industrial goods. *Journal of Marketing, 45 (4)*, 97-108.

Van Dierdonck, R., & Miller, J. G. (1980). Designing production planning and control systems. *Journal of Operations Management, 1(1)*, 37–46.

Varis, M., & Littunen, H. (2010). Types of innovation, sources of information and performance in entrepreneurial SMEs. *European Journal of Innovation Management, 13*, 128-154.

Verhoef, P. V., & Leeflang, P. S. H. (2009). Understanding the marketing department's influence within the firm. *Journal of Marketing, 73*, 14-37.

Voss, C. A. (2005). Alternative paradigms for manufacturing strategy. *International Journal of Operations & Production Management, 25 (12),* 1211–1222.

Ward, P. T., & Duray, R. (2000). Manufacturing strategy in context: Environment, competitive strategy and manufacturing strategy. *Journal of Operations Management, 18*, 123-138.

Wheelwright, S. C. (1984). Manufacturing strategy: Defining the missing link. *Strategic Management Journal, 5 (1)*, 77–87.

Wind, J. (1981). *Product lifecycle*. Unpublished manuscript. The Wharton School, University of Pennsylvania.

Wong, H. K., & Ellis, P. D. (2007). Is market orientation affected by the product lifecycle? *Journal of World Business, 42*, 145-156.

Youndt, M. A., Snell, S. A., Dean, J. W. Jr., & Lepak, D. P. (1996). Human resource management, manufacturing strategy and firm performance. *Academy of Management Journal, 39*, 836–866.

Zahra, S. A., & Covin, J. G. (1994). The financial implication of fit between competitive strategy and innovation types and sources. *The Journal of High Technology Management Research, 5 (2)*, 183-211.

Zeng, J., Phan, C. A., & Matsui, Y. (2015). The impact of hard and soft quality management on quality and innovation performance: An empirical study. *International Journal of Production Economics, 162*, 216-226.

Zeng, J., Zhang, W., Matsui, Y., & Zhao, X. (2017). The impact of organizational context on hard and soft quality management and innovation performance. *International Journal of Production Economics, 185*, 240-251.

DECISION SCIENCES INSTITUTE

An Exploration of the Relationship Between Lean Manufacturing Practices and Collaborative Product Development

(Full Paper Submission)

Mohammad Z. Meybodi
Indiana University Kokomo
mmeybodi@iuk.edu

ABSTRACT

The focus of this article is to explore potential relationship between lean manufacturing (LM) practices and collaborative product development (CPD). Analysis and comparison of several critical elements show high degree of resemblances between the two methods. A number of hypotheses regarding similarities between the elements of the two methods were developed and tested. Survey data from a sample of manufacturing organizations strongly supports the hypotheses regarding similarities between the two methods. Statistical results also indicate compared with conventional companies, by utilizing CPD, LM organizations are able to develop products faster with better quality, lower development cost, higher frequency, and lower manufacturing cost.

KEY WORDS: Collaborative, Product Development, Lean Manufacturing

INTRODUCTION

In a global market, innovation and speedy new product development is crucial for companies to gain competitive advantage. Creating new product ideas that are consistent with organizational strategy and moving these ideas through the stages of design, development, testing, and deployment has been the trade mark of successful world class organizations (Jacobs and Chase, 2020; Johansson and Safsten 2015; Ferioli et al. 2010; Roulet et al. 2010). Introducing new products to the market early has several strategic and tactical advantages. (Lofstrand, 2010; Kristav, 2016; Wen et.al. 2020; Cooper and Kleinschmidt, 1994; White, 1993).

Despite its well-known strategic role, for large number of manufacturing organizations innovation, design, and successful management of new product development has often been a major challenge. Long development time, prohibitive development and manufacturing costs, and questionable quality have been the common result for many of these organizations. The primary factor contributing to such unsuccessful result is perhaps the use conventional sequential method by these organizations (Morgan and Liker, 2006). However, manufacturing literature for the past three decades clearly shows that through their lean manufacturing practices, some world class organizations such as Toyota have dominated competition not only in the area of manufacturing but also in the area of innovation, design, development, and quick commercialization of new technologies (Marisa et al. 2008; Heinzen and Hoflinger 2017; Ulrich and Eppinger, 2004; Michael, 2008; Unger and Eppinger 2009). Instead of traditional sequential approach, world class organizations utilized collaborative product design and development (CPD) method. The focus of this article is to understand such contrast between the two type of organizations.

LITERATURE REVIEW

For the past two decades, LM has been a great force in the world of manufacturing. Some of the main benefits of a LM such as lower inventory, quicker delivery, and lower cost have been well documented (Cook and Rogowski, 1996; Hobbs, 1994; Temponi and Pandya, 1995; Deshpande and Golhar, 1995; Billesbach, 1991; Handfield, 1993; Lawrence and Hottenstein, 1995; Golhar, Stamm, and Smith, 1990; McKay, et.al 2011; Moras and Dieck, 1992). In the simplest form, LM requires maximizing value added production activities by removing unnecessary wastes. Identification and elimination of waste and respectful treatment of employee are the two fundamental principles of a LM system (Hobbs, 1994; Womack and Jones, 2003). Elimination of waste is achieved by adopting practices such as continuous quality improvement, setup time reduction, utilizing flexible resources, group technology layout, and pull production system (Gargeya, and Thompson, 1994; Sohal, Ramsay, and Samson, 1993). Respectful treatment of people often means employee empowerment; it includes elements such as team work, fair compensation, employee training and new positive attitude toward suppliers (Sohal, Ramsay, and Samson, 1993). Unfortunately, since its beginning in mid 1980's, often a narrow view of LM has been accepted and utilized by western manufacturers. Application of LM to reduce inventory and increase deliveries is only a small fraction of the full potential benefits of a LM system (Blackburn, 1991; Kristav, 2016). To take advantage of the full benefits of LM, one needs to have a much broader view of LM principles (Blackburn, 1991). Looking at LM as a process of eliminating waste and respectful treatment of employee, its principles can be applied to other areas including service areas such as healthcare, education, government, and new product development, (Womack and Jones (2003). Application of LM principles to new product development has great opportunity to shorten product development time, improve design quality, and reduce product development and manufacturing costs (Anand and Kodali, 2008). The company that originated famous LM system, TPS, also developed Toyota Product Development System (TPDS). TPDS employs LM principles and tools such as value stream mapping, Kanban, 5S system, and continuous improvement to eliminate waste from product development activities and bring quality products to market faster than their leading competition (Morgan and Liker, 2006; Ward, 2007). However, TPDS is a comprehensive strategy that involves various approaches to eliminate waste from new product development activities. The focus of this article is on special case of TPDS. The objective of the article is to answer the following questions:

1. Are there similarities between LM and CPD practices?
2. Are there differences between new product development performances for LM companies using CPD and conventional companies?

The remainder of the article is organized in the following manner: First, an overview of the differences between conventional sequential method and recent CPD is presented. Second, the article compares and analyzes similarities between LM and CPD for a number of critical elements followed by a set of test of hypotheses on similarities between the elements. Third, the article tests product design performance for conventional sequential method and CPD method. Research methodology, results, and conclusion are the final sections of the article.

CONVENTIONAL AND COLLABORATIVE METHODS OF PRODUCT DEVELOPMENT

For the past two decades, LM has been a great force in the world of manufacturing. Some of the main benefits of a LM such as lower inventory, quicker delivery, and lower cost have been well documented (Cook and Rogowski, 1996; Hobbs, 1994; Temponi and Pandya, 1995; Deshpande and Golhar, 1995; Billesbach, 1991; Handfield, 1993; Lawrence and Hottenstein, 1995; Golhar, Stamm, and Smith, 1990; McKay, et.al 2011; Moras and Dieck, 1992). In the simplest form, LM requires maximizing value added production activities by removing unnecessary wastes. Identification and elimination of waste and respectful treatment of employee are the two

fundamental principles of a LM system (Hobbs, 1994; Womack and Jones, 2003). Elimination of waste is achieved by adopting practices such as continuous quality improvement, setup time reduction, utilizing flexible resources, group technology layout, and pull production system (Gargeya, and Thompson, 1994; Sohal, Ramsay, and Samson, 1993). Respectful treatment of people often means employee empowerment; it includes elements such as team work, fair compensation, employee training and new positive attitude toward suppliers (Sohal, Ramsay, and Samson, 1993). Unfortunately, since its beginning in mid 1980's, often a narrow view of LM has been accepted and utilized by western manufacturers. Application of LM to reduce inventory and increase deliveries is only a small fraction of the full potential benefits of a LM system (Blackburn, 1991; Kristav, 2016). To take advantage of the full benefits of LM, one needs to have a much broader view of LM principles (Blackburn, 1991). Looking at LM as a process of eliminating waste and respectful treatment of employee, its principles can be applied to other areas including service areas such as healthcare, education, government, and new product development, (Womack and Jones (2003). Application of LM principles to new product development has great opportunity to shorten product development time, improve design quality, and reduce product development and manufacturing costs (Anand and Kodali, 2008). The company that originated famous LM system, TPS, also developed Toyota Product Development System (TPDS). TPDS employs LM principles and tools such as value stream mapping, Kanban, 5S system, and continuous improvement to eliminate waste from product development activities and bring quality products to market faster than their leading competition (Morgan and Liker, 2006; Ward, 2007). However, TPDS is a comprehensive strategy that involves various approaches to eliminate waste from new product development activities. The focus of this article is on special case of TPDS. The objective of the article is to answer the following questions:

1. Are there similarities between LM and CPD practices?
2. Are there differences between new product development performances for LM companies using CPD and conventional companies?

The remainder of the article is organized in the following manner: First, an overview of the differences between conventional sequential method and recent CPD is presented. Second, the article compares and analyzes similarities between LM and CPD for a number of critical elements followed by a set of test of hypotheses on similarities between the elements. Third, the article tests product design performance for conventional sequential method and CPD method. Research methodology, results, and conclusion are the final sections of the article.

COMPARISON OF LEAN MANUFACTURING AND CPD ELEMENTS

For the past three decades, there has been an extensive volume of research in the area of LM. As a result, there is a set of generally accepted guidelines that organizations can follow to achieve manufacturing success. However, there has been limited research on the application of LM practices to product development and there is no comparable set of guidelines for successful management of the process. Recently, a number of world class PD companies have attempted to apply the principles of LM to PD activities (De Waal and Knott, 2019). The company that started the most famous LM system, Toyota Production System (TPS) is also started Toyota Product Development System (TPDS). TPDS employs LM principles and enable the company to bring the highest quality products to market faster than their leading competition. Also, a number research on the application of LM principles to PD process has shown that achieving certain manufacturing process improvement such as reducing variation, reducing rework and yield loss, solving process bottlenecks, and managing capacity, can significantly reduce PD times.

Similarities between LM and CPD for a number of critical elements are shown in Table 1, (Goffin et. al, 2019; Spencer and Guide, 1995). Following is a brief comparison and analysis of selected elements in Table 1:

Layout

Layout in LM environment is often in the form of product focus and manufacturing cells. This type of layout is necessary because small lot size production requires that the layout to be compact and efficient to ensure smooth flow of materials and close communication between work stations. Unlike conventional manufacturing, where material is pushed forward, the flow in a LM environment is in two directions; material is pulled forward, but information flows backward to provide feedback on performance and material requirements.
In CPD, overlapping of a large number of activities requires a complete change in layout that facilitates communication and encourages team work. Instead of organizing by sequential functions, CPD emphasizes on cross-functional integration and the formation of a design team. The design team sits together in one location, creating a type of project layout. A project layout creates an environment for frequent, two-way communication between team members, which encourages concurrent development of a product and its associated processes.

Lot Size

In contrast to conventional manufacturing, LM manufacturing requires production of small lot-sizes. Production of small lot-sizes is possible by drastically reducing set-up times. It is well documented that production of small lot-sizes in LM is closely associated with improved quality, reduced inventory, faster delivery, and more responsive to market demands.
Similar to LM, CPD also utilizes small lot-sizes; the only difference is that in LM small lot sizes of goods are processed but CPD requires small lot-sizes of information. That is, continuous two way communication in CPD is similar to early release of small batches of information (Blackburn, 1991; White, 1993). With the early release of small batches of information, downstream constituents can begin working on different phases of the design while final design is evolving. The early release of information reduces uncertainty and encourages early detection of problems, which enables organizations to avoid costly, time-consuming changes.

Employee and Supplier Involvement

In LM environment, management encourages employee involvement and team work. The responsibility for job scheduling and quality are often passed to the teams at the shop floor. Due to small lot size production, delegation of authority to the teams at the shop floor is essential for smooth production flow. Also, in LM suppliers work closely with manufacturing organization to improve quality and shorten delivery time.
Similar to LM, in CPD the responsibility for scheduling of the activities pushed down to product development team at the lowest level. Passing responsibility down to the team is essential to achieve a high level of activity coordination and information sharing among team members. Also, in CPD suppliers work closely with the design team to reduce development costs, shorten development time, and offer ideas toward improving the quality of the design.

Quality

In LM and CPD environments, organizations are often proactive and quality means getting it right the first time. In LM, since batch sizes are small quality at the source and continuous quality improvement are the main foundations. Shop floor workers are empowered to become their own inspectors responsible for the quality of their output. In CPD, because of the teamwork and two-way flow of information between team members, and utilization of quality improvement tool such as six sigma process quality problems are detected earlier and solved before they have a cumulative impact on the rest of the project (Chakravorty and Franza, 2009).

Technology

In a LM system, technology is not viewed as a substitute, or shortcut to process improvement. Rather, technology has been utilized after process analysis and simplification has been performed. The role of technology in CPD is also enormous; it requires that the design team with diverse expertise makes a large number of interrelated decisions regarding the form, fit, function, cost, quality, and other aspects of the design (Karagozoglu and Brown, 1993). This requires supply and processing of relevant information from multiple sources in a coordinated manner. Successful organizations use technology in their PD process similarly to the way they use technology in their LM system. In CPD, the design team utilizes appropriate technologies and tools at various stages of PD process. Effective use of technologies and tools can dramatically shorten PD time, reduce number of prototypes, cut costs, and improve quality of the design (McKay et al. 2011; Yamamoto and Abu Qudiri 2011; Roulet et.al (2010). The key to the success of technology in CPD is building an effective design team with open cross-functional communication lines.

FACTOR HYPOTHESES

Comparison and analysis of elements in Table 1 show a high degree of similarities between LM and CPD. To study further, a set of twenty five hypotheses (H1-H25) that statistically test similarities between LM and CPD will be presented. The hypotheses are shown in Table 2. Each hypothesis in Table 2 consists of two parts- a and b. In part a, the test is conducted for LM elements and the corresponding test for CPD elements is conducted in part b. The last hypothesis examines the overall impact of LM principles on CPD.

Hypotheses (H1-H25):

There is a high degree of similarities between lean manufacturing and CPD elements.

PRDUCT DEVELOPMENT PERFORMANCES

The following dimensions of quality, time, competency, development cost, and manufacturing cost are used to measure the performance of NPD (Ulrich and Eppinger, 2000; Wheelwright and Clark, 1992):

- **Quality**: Quality is ultimately reflected in the price customers are willing to pay, the market share, and the bottom line profit. In PD, quality problems are often the results of incomplete information and miscommunication among various functions. Quality often means a minimal number of redesign or rework. In this article, number of design changes during the development process and early manufacturing phase is used as a measure of design quality.

- **Development time**: Development time is the length of time between initial idea generation until new product is ready for introduction to the market. Shorter development time raises the competitive value of new product in terms of premium price, larger market share, and higher profit margin.

- **Development competency**: Development competency is the ability of the organization to develop future products better, faster, and cheaper. Competent workforce and effective use of technologies are important elements of organizational PD competency. Frequency of new product introduction to the market is used as a measure of development competency.

- **Development cost**: This is the total cost from the early idea generation until the product is ready for manufacturing. For most organizations, development cost is usually a significant portion of the budget and must be considered in light of budget realities and the timing of budget allocations.

- **Manufacturing cost**: Manufacturing cost includes initial investment on equipments and tools as well as the incremental cost of manufacturing the product. There is a close relationship between manufacturing cost and the type of decisions made during the early design stage. Although early design decisions determine about 70 percent of future manufacturing cost, organizations often spend far too little time and resources during this stage (Huthwaite, B. 1991). To save future manufacturing cost, it is prudent for the companies to spend more time and resources during the early design phases of PD process where critical design decisions are made.

PERFORMANCE HYPOTHESES

In the second set of hypotheses (H26-H30), the differences between PD performances for LM manufacturing companies and conventional companies are tested.

Hypotheses (H26-H30):
H26: By utilizing CPD approach, LM companies are able to design new products with fewer design changes than conventional companies(better quality).

H27: By utilizing CPD approach, LM companies are able to design new products faster than conventional companies.

H28: By utilizing CPD approach, LM companies are able to design new products more often than conventional companies.

H29: By utilizing CPD approach, LM companies are able to design new products with less development cost than conventional companies.

H30: By utilizing CPD approach, LM companies are able to design new products with less manufacturing cost than conventional companies.

RESEARCH METHODOLOGY

The target population for this study consisted of manufacturing firms in the states of Illinois, Indiana, Ohio, Michigan, and Wisconsin. A sample of manufacturing firms with more than 50 employees was chosen from manufacturers' directories of those states. The sample covers organizations in variety of industries ranging from fabricated metal, communication, electronics, automotive, toots, chemicals, rubber, and paper products. A comprehensive survey instrument based on examination of the literature and critical elements listed in Table 1 was developed. A panel of practitioners and researchers with experience in LM and NPD was used to validate the survey. Cronbach alpha reliability test was also used to validate the survey. The minimum Alpha value for all factors was 0.72. In addition to general organization and managerial profile items, the survey contained 50 items (25 paired) regarding similarities between LM and CPD elements. The twenty five paired questionnaire items are shown in Table 2.

Also, the survey instrument contained a number of questionnaire items on PD performances for LM companies using CPD and conventional companies. Out of 91 completed surveys received, 84 surveys were usable resulting in a response rate of 17%. Based on a number of questionnaire items on the principles of LM practices, 33 organizations were grouped as LM companies and 51 organizations were categorized as conventional companies.

The survey data indicates that majority of respondents had various high level managerial positions from 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 and PD experience, about 28% of the respondents had between 10 to 20 years and 60% had more than 20 years of manufacturing experience. About 72% of the sample had more than 10 years of LM experience and close to 65% of the sample had more than 10 years of PD experience.

RESEARCH RESULTS

As stated earlier, in the first set of hypotheses the objective was to examine similarities between LM and CPD for a set of paired elements shown in Table 2. For each item, the null hypothesis was that the mean response for LM is equal to the mean response for CPD. The differences between the mean responses for LM and CPD were compared using two independent populations statistical t-test. The respondents were asked to rate each element of Table 2 based on the degree of their agreement on Likert-type scale of 1 to 5 to the questions, where (1=strongly disagree; 2= disagree; 3 = indifferent; 4=agree; and 5=strongly agree). Table 3 shows the result of similarities between LM and CPD.

As shown in Table 3, overall the respondents strongly agreed with the statements regarding similarities between LM and CPD elements. The mean ratings for about 70% of the elements for both LM and CPD are above 3.80. Specifically, out of twenty hypotheses, the respondents agreed that there is a high degree of similarities between LM and CPD for all except three hypotheses H4, H7, and H9.

For H4, the mean ratings for LM and CPD are respectively 4.34 and 3.81. This means although the respondents understood that short set-up and fast transition time are the main requirements of successful LM and CPD, the relationship between short set-up and LM was much stronger. This is a reasonable result because an average manufacturing manager has longer experience with LM than CPD. They clearly understood that successful LM requires small lot-size and small lot-size requires short set-up time. However, due to their shorter experience with CPD and because CPD is primarily an information processing process, the links between small batches of information and fast transition time is not clear. H7 hypothesizes the relationships between small lot-sizes and quality improvement for both LM and CPD. For this test, the mean ratings for LM and CPD are respectively 3.43 and 3.89. This indicates for an average manager it is easier to recognize the relationship between CPD and quality improvement than the relationship between LM and quality improvement. The higher rating for CPD is perhaps due to continuous and two way communication among design team members, which encourages early detection of the design problem. The LM result is also consistent with the literature because although total quality management and quality improvement are fundamental requirements of successful LM, an average manufacturing manager has difficulty to understand this relationship. The relationships between small lot-size and reduced manufacturing cost in LM and the relationship between small batches of information and reduced development cost in CPD are examined in H9. The mean ratings for LM and CPD are respectively 3.58 and 3.94. For the same reasons as H7, this means for an average manager it is easier to understand this relationship in CPD than

LM. The LM result is interesting and also consistent with the literature because reduced manufacturing cost in LM is primarily due to elimination of wastes, a fundamental principle of LM, and an average manufacturing manager has difficulty to see this relationship. The overall impact of LM principles on LM and CPD is examined in H24. It is obvious that the data supports the hypothesis as the mean ratings for LM and CPD are respectively 4.56 and 4.29 indicating strong agreement with the statements that the main principles of waste elimination and respectful treatment of people in LM can also be applied in CPD.The last column of Table 3 shows correlation coefficients between LM and corresponding CPD elements. The correlation coefficients in Table 3 strongly support the above analysis. With the exception of three hypotheses H4, H7, and H9 other coefficients are greater than 0.60 indicating a high degree of linear association between LM and CPD elements.
The performance hypotheses (H26-H30) state that by utilizing CPD approach, LM companies are able to design new products with fewer design changes, faster, more often, with less development cost, and less manufacturing cost than conventional companies.

Table 4 provides useful statistical information regarding PD performances for LM and conventional companies. The average number of design changes for conventional and LM companies are respectively 5.36 and 3.28, a quality improvement of 63%. The average development time for conventional and LM companies are respectively 37.52 and 24.73 months, an improvement of 52%. For development competency, the average time between introduction of new products for conventional companies is 49.46 months and 32.72 months for LM companies, an improvement of 51%. Table IV also indicates that LM organizations enjoy a 45% reduction in PD cost and 36% reduction in manufacturing cost. From the last column of Table 4, it is clear that the hypotheses are strongly supported by the data as the p-value for all five hypotheses is less than 0.005.

CONCLUSION

The focus of this article was to demonstrate possible links between LM practices and CPD. First, comparison and analysis of a number of elements showed remarkable similarities between LM practices and CPD. Second, a set of paired hypotheses was used to test similarities between LM practices and CPD elements. Statistical results clearly support the hypotheses regarding similarities between LM and CPD for majority of elements. Specifically, out of twenty four hypotheses, the respondents agreed that there is a high degree of similarities between LM and CPD for all but three hypotheses. The last pair of hypotheses that examines the overall impact of LM principles is especially important. Statistical results strongly agreed that the main principles of waste elimination and respectful treatment of people in LM is also applicable to CPD. The correlation coefficients between LM and CPD elements also supported the same result. Third, statistical results also indicate that compared with conventional companies, LM companies are able to develop new products with 63% better quality, 52% less development time, 45% less development cost, and 36% less manufacturing cost. Also frequency of new product introduction is 51% faster than conventional companies.

In summary, statistical results of this article clearly show strong links between LM practices and CPD. Managerial implication of the research is that successful implementation of LM principles goes much beyond inventory reduction and frequent deliveries. For LM organizations success in CPD is the result of knowledge and technology transfer from their LM system into their NPD process.

(Tables and references are available from the author upon request)

DECISION SCIENCES INSTITUTE

A Uniform-Price Auction Mechanism of Carbon Allowance with Flexible Supply and Asymmetry Buyers

Zijuan Hu

School of economics and management, Tongji University

Email: zijuan_hu@163.com

Jie Lin

Email: jielinfd@163.com

School of economics and management, Tongji University

ABSTRACT

To tackle the problems of arbitrary underpricing equilibrium and "invisible collusion" in the uniform-price auction with a fixed supply, a flexible-supply auction mechanism for carbon allowance is proposed. Regulated and non-regulated bidders are allowed to participate simultaneously, and the optimal supply of regulated bidders is determined after the seller having observed the bids. For this mechanism, we explore the formation process of asymmetric linear and nonlinear Nash equilibrium of regulated bidders and further discuss the equilibrium price and the optimal supply in each equilibrium. It concludes that the proposed auction mechanism satisfies the participation and incentive compatibility constraints. Meanwhile, there exist multiple equilibriums, but they can effectively eliminate implicit collusion and underpricing equilibrium. Finally, suggestions are drawn about the mechanism design of carbon allowance auctions from market access threshold, initial bidding quantity, and bidder's bidding strategy aspects.

KEYWORD: Carbon emission right, Elastic supply, Uniform-price auction, Asymmetric bidders, Non-regulated bidders

INTRODUCTION

To cope with the global environmental crisis, the Cap-and-trade emission trading scheme (ETS) has become one of the most effective mitigation measures. In 2005, the European Union implemented the first cap-and-trade carbon emission trading (EUETS) system. Since then, the Tokyo Emissions Trading System, the California Emission Trading System, and the Quebec Emissions Trading System have been successively established between 2005 and 2013(Li 2015; Newell 2013). As the world's largest emitter of carbon dioxide, China has successively launched seven regional carbon allowances trading pilots, including Beijing, Tianjin, Shanghai, Chongqing, Guangdong, Fujian, and Shenzhen since 2013. China has launched a national carbon emissions trading market at the end of 2017, which promises to reduce CO_2 emissions per unit of GDP by 40%-50% of that in 2005 by 2020(Cui 2014).
The nationwide carbon emissions trading market covers two levels of markets: the primary market allocates the initial carbon allowance, and the secondary market is the trading market. As the basis and critical link of the carbon allowances trading system, the allocation of initial carbon allowance is a crucial area of government concern and academic research. According to the current ETS, the most common initial carbon allowance distribution methods include grandfathering (firms are given carbon allowance for free), auction, and combining these two forms (Hitzemann 2015, Zhang 2015). An auction can not only overcome the drawbacks of the

grandfathering, such as “the whipping of fast cows, unfair punishment”(Anderson 2011), but it also generates the so-called “double dividend effect”(Parry 1999), which means auction can not only achieve the optimal allocation of carbon allowance but also provides financial support for environmental protection projects. Therefore, the carbon allowance auction has attracted much attention from all over the world.

LITERATURE REVIEW

In terms of participants in the carbon allowance auction, many studies have found that the existence of financial institutions or individual investors can play a positive role in the liquidity and effectiveness of the carbon market. Liu et al. (2015) believe that the active participation of financial institutions in the carbon market will promote effective emissions reduction and attract investors to participate in carbon market transactions. Through experimental research on non-regulated companies, Chesney(2011) and Taschini (2014) simulated the allowance trading process of the European carbon market. They found that allowing non-regulated companies to participate can promote market activity and help regulated companies achieve compliance faster. However, in the existing literature, only a few studies consider the existence of non-regulated buyers in the construction of carbon trading models, which is not conducive to the in-depth research of auction mechanism design. Therefore, we propose a uniform-price auction with elastic supply volume that considers the simultaneous participation of regulated and non-regulated enterprises. On the one hand, it can solve the problem of low-price equilibrium caused by a fixed supply, and it is also suitable to strengthen the liquidity of the auction market.

Besides, most existing studies on emission auctions are based on the buyer's information symmetry or only consider linear equilibrium (Rao 2012b). In the real auction market, the valuation of carbon emission right per unit is directly related to the cost of emission reduction. The heterogeneous energy consumption structure and emission efficiency of the emitters lead to great differences in the cost of emission reduction among the regulated firms, which belongs to the private information of bidders and is asymmetric.

Carbon Emission Right Auction Methods

In the existing research on carbon emission right auction, the comparison of carbon emission right auction methods has aroused heated discussion among scholars. Cramton (1998) points out that static auctions are easy to execute and have low transaction costs, avoiding the collusion of bidders in multiple auctions. In static auctions, the most common pricing methods include uniform and discriminatory price auctions. Theorists have proved that the uniform-price auction has more advantages in the auction of carbon emission rights from the aspects of government revenue, fairness, and economic impact. Wilson (1979) compares the two auction modes. He finds that the government revenue under discriminatory price auctions is higher than that under uniform price auction. Hence, the cost of performance of enterprises under discriminatory price auction is higher than uniform-price auction. Hu et al. (2019) carried out simulation experiments. They found that the uniform-price auction is better when the relationship between market supply and demand is unclear, considering enterprises' cost and fairness. In practice, the uniform-price auction has been more widely promoted in carbon emission rights auction because of its simplicity and practicability. Therefore, we focus on the uniform-price auction mechanism design of carbon allowance in this paper.

In the design of the uniform-price auction mechanism, the existing researches are rich in market power (Ausubel 1998; Min 2016), risk attitude (Wang 2002; Gretschko 2015), commission constraints (Wang 2004; Zhou 2012). Most relavent studies about the carbon emission rights auction are based on fixed supply ($\sum_{i=1}^{n} q_i = Q_0$), which is prone to low-price equilibrium and "invisible collusion" (Rao 2012a) and then leads to inefficient auction results. Therefore, some scholars have begun to pay attention to the design of the auction mechanism with an undetermined total amount. Taking the substitutability between commodities into account, Fan et al. (2005) determined the auction amount of each product in the multi-item auction by maximizing all the lots' overall profit and reducing the external loss of the substitute goods. Rao et al. (2012a) proposed an auction mechanism based on the symmetry bidders and the coexistence of three risk attitudes, in which the uncertain number of auctions depends on the transaction price.

This paper proposes a uniform-price carbon emission right auction model with a flexible supply by introducing non-regulated (without mandatory emission reduction requirements) bidders into the auction model. To make the model more general, this paper extends all existing assumptions of buyer symmetry to the case of buyer asymmetry. It calculates the linear and nonlinear equilibrium quotation of asymmetric buyers. It is verified that the auction method satisfies the constraints of participation and incentive compatibility. Finally, further suggestions on the design of the auction market mechanism of government carbon emission rights are put forward.

The remainder of the paper is organized as follows. Section 2 gives some assumptions and presents a uniform-price auction mechanism based on flexible supply and asymmetric bidders. Section 3 solves the Nash equilibrium strategy based on linear and nonlinear equilibrium bids, respectively. Section 4 further analyses the equilibrium strategy and give some extended suggestions on auction mechanism design of carbon emission right. Section 5 concludes the paper.

MODEL

Assumptions and symbolic descriptions:
The theoretical assumptions of the common homogeneous multi-item uniform-price auction are as follows (Wilson 1979; Klemperer 1989; Rao 2012):

H1: There are an auctioneer and n buyers (n>2).

H2: Both auctioneer and buyers are risk-neutral.

H3: The total amount of the auction is Q_0.

H4: All buyers have a common valuation v of auction goods.

H5: Each bidder i (i=1,2,3,…,n)submits a non-increasing and continuous demand function $d_i(p)$ at the price of p, and aggregate demand function $D(p) = \sum_{i=1}^{n} d_i(p)$.

H6: The auctioneer uniformly distributes goods under the clearing price p_0, and the number of goods allocated to the buyer i is q_i, and $\sum_{i=1}^{n} q_i = Q_0$.

Based on the above assumptions H1-H6, the purpose of the auctioneer is to optimize his auction profits, which is $\max \Pi = p_0 \sum_{i=1}^{n} q_i = p_0 Q_0$. The bidders' goal is to determine the appropriate bidding strategies to maximize the auction surplus $\max U_i = (v - p_0) q_i$ (i=1,2,3,…,n). Based on the classic model, we use the initial allocation of carbon emission rights as the design background of the auction mechanism and further expands the assumptions:

H1 is improved to $H'1$: Bidders include two categories. The first category is m ($m \geq 2$) competitive regulated companies with mandatory emission reduction pressures (hereinafter referred to as regulated bidders), and the second category is financial institution investors and individual investors, who have no pressure to reduce emissions (hereinafter referred to as non-regulated bidders), who have increased carbon market liquidity.

H3 is improved to $H'3$: The total amount of carbon emission right allocated to all regulated buyers is uQ_0 ($0<u<1$), and the amount of carbon emission right obtained by the non-regulated buyers is not exceeded Q_1 ($0 \leq Q_1 \leq Q_0$).

H4 is improved to $H'4$： Assuming that the buyers are independent of each other, the abatement cost of the regulated enterprise i can be expressed as $m_i(q_i)=z_i + a'q_i$, $i=1,2,3..m$, where z_i represents the fixed abatement cost of the regulated enterprise i, and a' is the marginal abatement cost. On the contrary, the main purpose of non-regulated buyers to participate in the auction market is “Purchasing at a low price in the auction market and selling at a high price in the secondary market.” Therefore, it is assumed that the non-regulated buyers in the auction market have the same maximum value of the unit of carbon emission rights p_1 , which can be understood as the average price of the secondary market carbon allowance. Since the carbon price in the secondary market will not be higher than the marginal abatement cost of the enterprise, there is $0 < p_1 \leq a'$， $i=1,2,3...m$.

The government issues an auction announcement, and each regulated buyer i submits a continuous and divisible non-increasing quotation strategy $d_i(p)$. After observing the quotation strategies of all the regulated buyers $\Omega=\{d_1(p), d_2(p), . . . , d_m(p)\}$, the auctioneer determines the elastic allocation coefficient u according to $\sum_{i=1}^{m} q_i = uQ_0$. If $u=1$, the emission rights are allocated to the regulated buyers, otherwise $u=0$, all carbon allowances are allocated to the non-regulated buyers. In other words, the carbon emission right allocated to the regulated buyer is no longer a fixed amount, which is dependent on the elastic supply of the distribution coefficient u ($u \in (0,1)$).

It is assumed that the government adopts uniform-price auctions to allocate the initial carbon emission right. The government aims to design the auction pricing and distribution rules to obtain the maximum auction profits. The model is expressed as follows:

$$\max \Pi = p(u)Q_0$$

$$\text{Subject to} \begin{cases} \sum_{i=1}^{m} q_i = uQ_0 \\ d_i[p(u)] = q_i \quad i=1,2,...,m \\ 0 < p(u) \leq p_1 \\ 0 \leq (1-u)Q_0 \leq Q_1 \\ 0 \leq Q_1 \leq Q_0 \end{cases}$$

where $0 < p(u) \leq p_1$ represents the upper and lower bounds of the equilibrium price, not higher than the highest bidding price of the non-regulated buyers. We don't formulate the situation of $p(u) > p_1$, because at this time, all carbon allowances are only allocated to the regulated buyers, which is consistent with the proposed uniform-price auction mechanism by Wang (2016). However, we mainly formulate the situation in which the regulated and non-regulated buyers are involved in the auction process. The latter three constraints are the limit on the maximum amount of carbon allowance allocated to non-regulated buyers.

Assuming that all regulated buyers' valuation of carbon allowances is based on their marginal abatement costs. The goal is to maximize their earnings by selecting appropriate quotation strategies. The profit of regulated buyers R_i can be expressed as savings in abatement costs and purchase costs.

$$\max R_i = z_i + a' q_i - pq_i \tag{1}$$

The market-clearing price can be defined as the lowest price in the winning bidders.

$$p_k = \inf\{p \mid D(p) \geq uQ_0\} \tag{2}$$

The auction process based on elastic supply volume discussed above is as follows：

(1)The government announces auctions amount Q_0 and the auction rules, both regulated and non-regulated firms have access to the auction；

(2) Each bidder *i* submits a non-increasing and continuous demand function $d_i(p)$；

(3) The government selects the optimal elastic distribution coefficient u* according to the set of quotes of all the regulated buyers $\Omega=\{d_1(p), d_2(p), \ldots, d_m(p)\}$, and then determines the equilibrium price to achieve the goal of maximizing auction profits;

(4) The government allocates carbon allowance to all buyers at a uniform price p_u, where the carbon allowance allocated to regulated buyer *i* is q_i, and the total allowance to non-regulated emission buyer is $(1-u)Q_0$.

EQUILIBRIUM ANALYSIS

Definition 1. Let $\Omega=\{d_1(p), d_2(p), \ldots, d_m(p)\}$ be the bidding set of all regulated buyers. If $R_i(d_i^*(p), d_{-i}^*(p)) \geq R_i(d_i(p), d_{-i}^*(p))$ is holds for any $i \in \{1,2....m\}$ and $d_i(p) \in D_i(p)$, then （$d_1^*(p)$, $d_2^*(p),...,d_i^*(p)$) is called an equilibrium bidding strategy, where $d_{-i}(p)$ represents a set of bidding strategies other than *i*, that is $\{d_1(p), d_2(p),..., d_{i-1}(p), d_{i+1}(p), d_m(p)\}$.

According to the definition of Nash equilibrium, if ($d_1^*(p)$, $d_2^*(p),...,d_i^*(p)$) is the Nash equilibrium of the auction model, then for any regulated buyer *i*, the equilibrium price under the uniform-price auction satisfies $p(u)^* = \arg\max_p\{R_i = z_i + (a'-p)(uQ_0 - \sum_{\substack{j=1 \\ j \neq i}}^{m} q_j)\}$, while satisfying the market clearing condition $\sum_{i=1}^{m} q_i(p^*) = uQ_0$.

3.1 Linear equilibrium strategy

Proposition 1. Suppose regulated buyers are asymmetric and risk-neutral, the linear equilibrium quotation strategy for regulated buyer *i* is.

$$d_i(p)=\overline{q_i}-\frac{(m-1)\overline{Q}-(m-2)uQ_0}{a^{'}(m-1)[muQ_0-(m-1)\overline{Q}]}[uQ_0-(m-1)\overline{q_i}]p \tag{3}$$

where $\overline{q_i}$ is the demand for regulated bidders at the reserve price of 0, and $\overline{Q}=\sum_{i=1}^{m}\overline{q_i}$ stands for the maximum demand of all regulated bidders $i\in\{1,2....m\}$. When $\overline{q_i}\geq\frac{uQ_0}{m}$ is satisfied, there will be no excess supply of carbon allowances in the auction market.

Proof. It is assumed that the quotation strategy of the regulated buyer *i* is $d_i(p)=a_i-b_ip$,($a_i>0,b_i>0$), and when the auction reserve price $p=0$, the demand of the regulated bidder *i* is $\overline{q_i}$.

$$d_i(0)=a_i=\overline{q_i} \tag{4}$$

When the equilibrium is reached, the demand for regulated buyers is satisfied with

$$\sum_{i=1}^{m}q_i(p^*)=uQ_0 \tag{5}$$

Let the bidder's quotation other than *i* be equal to $\sum_{\substack{j=1\\j\neq i}}^{m}q_j(p^*)$, and then the optimal response strategy $y_i(p)$ satisfies with

$$y_i(p)=uQ_0-\sum_{\substack{j=1\\j\neq i}}^{m}q_j(p^*) \tag{6}$$

The benefit of regulated buyer *i* in Eq.(1) can be expressed as

$$R_i=z_i+(a'-p)(uQ_0-\sum_{\substack{j=1\\j\neq i}}^{m}q_j) \tag{7}$$

According to the first-order condition of maximizing the benefits $\frac{dR_i}{dp}=0$ and $q_i=a_i-b_ip$, then

$$q_i=(a'-p^*)\sum_{\substack{j=1\\j\neq i}}^{m}b_j \tag{8}$$

From (8) to (5), we have

$$\sum_{i=1}^{m}b_i=\frac{uQ_0}{(a'-p^*)(m-1)} \tag{9}$$

And because $\sum_{i=1}^{m}b_ip^*=\overline{Q}-uQ_0$, then bring it into (9)

$$p^*=a'[1-\frac{uQ_0}{(m-1)\overline{Q}-(m-2)uQ_0}] \tag{10}$$

Since $\frac{d^2R_i}{dp^2}$=-2$\sum_{\substack{j=1\\j\neq i}}^{m}b_j<0$, the second-order condition of maximization is satisfied.

From(10)to(9), we obtain

$$\sum_{i=1}^{m}b_i=\frac{1}{a'}(\overline{Q}-uQ_0+\frac{uQ_0}{m-1}) \tag{11}$$

Let Eq.(11)minus Eq.(8), we have

$$\sum_{i=1}^{m} b_i - \sum_{\substack{j=1 \\ j\neq i}}^{m} b_j = \frac{1}{a'}(\overline{Q} - uQ_0 + \frac{uQ_0}{m-1}) - \frac{\overline{q_i} - b_i p}{(a' - p^*)} \tag{12}$$

Substituting(12)into(10), and we get b_i

$$b_i = \frac{(m-1)\overline{Q} - (m-2)uQ_0}{a'(m-1)[muQ_0 - (m-1)\overline{Q}]}[uQ_0 - (m-1)\overline{q_i}] \tag{13}$$

Replace a_i and b_i in $d_i(p) = a_i - b_i p$, then

$$d_i(p) = \overline{q_i} - \frac{(m-1)\overline{Q} - (m-2)uQ_0}{a'(m-1)[muQ_0 - (m-1)\overline{Q}]}[uQ_0 - (m-1)\overline{q_i}]p \tag{14}$$

Proposition 2. When the market reaches equilibrium, the government gives the buyer the optimal elastic distribution coefficient u^* and p^* satisfies the following formula.

$$u^* = \max\{\frac{Q_0 - Q_1}{Q_0}, \frac{(a' - p_1)(m-1)}{(a' - p_1)(m-2)+1} \times \frac{\overline{Q}}{Q_0}\} \tag{15}$$

$$p^* = \min\{p_1, a' - \frac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)}\} \tag{16}$$

Proof. According to Eq.(10)

$$p^*(u) = a' - \frac{uQ_0}{(m-1)\overline{Q} - (m-2)uQ_0}$$

$$\frac{dp^*}{du} = \frac{-(m-1)Q_0\overline{Q}}{[(m-1)\overline{Q} - (m-2)uQ_0]^2} < 0 \tag{17}$$

$p^*(u)$ is the monotonic decreasing function of u .

As is assumed above, we have $p(u) \le p_1$, when $p^*(u) = p_1$

$$u^* = \frac{(a' - p_1)(m-1)}{(a' - p_1)(m-2)+1}\frac{\overline{Q}}{Q_0} \tag{18}$$

Because u needs to be satisfied with $(1-u)Q_0 \le Q_1$, that is $u \ge \frac{Q_0 - Q_1}{Q_0}$, and then

$$u^* = \frac{(a' - p_1)(m-1)}{(a' - p_1)(m-2)+1} \times \frac{\overline{Q}}{Q_0} \ge \frac{Q_0 - Q_1}{Q_0} \tag{19}$$

So we have $p_1 \le a' - \frac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)}$, if $p_1 \le a' - \frac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)}$ is true, equilibrium price $p^*(u) = p_1$, At this time, the optimal elastic distribution coefficient $u^* = \frac{(a' - p_1)(m-1)}{(a' - p_1)(m-2)+1} \times \frac{\overline{Q}}{Q_0}$; if $p_1 > a' - \frac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)}$ is true, $u^* = \frac{(a' - p_1)(m-1)}{(a' - p_1)(m-2)+1} \times \frac{\overline{Q}}{Q_0} < \frac{Q_0 - Q_1}{Q_0}$, $p^*(u) < p_1$. Since $p^*(u)$ is the monotonic decreasing function of u , and $u \ge \frac{Q_0 - Q_1}{Q_0}$, so when $u = \frac{Q_0 - Q_1}{Q_0}$, the equilibrium price $p^*(u)$ gets the maximum value

$$p^* = a' - \frac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)} \tag{20}$$

$$p^* = \begin{cases} a' - \dfrac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)}, & p_1 > a' - \dfrac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)} \\ p_1, & p_1 \le a' - \dfrac{Q_0 - Q_1}{(m-1)\overline{Q} - (m-2)(Q_0 - Q_1)} \end{cases}$$

$$u^* = \begin{cases} \dfrac{(a'-p_1)(m-1)}{(a'-p_1)(m-2)+1} \times \dfrac{\overline{Q}}{Q_0}, & \dfrac{(a'-p_1)(m-1)}{(a'-p_1)(m-2)+1} \times \dfrac{\overline{Q}}{Q_0} \ge \dfrac{Q_0 - Q_1}{Q_0} \\ \dfrac{Q_0 - Q_1}{Q_0}, & \dfrac{(a'-p_1)(m-1)}{(a'-p_1)(m-2)+1} \times \dfrac{\overline{Q}}{Q_0} < \dfrac{Q_0 - Q_1}{Q_0} \end{cases}$$

3.2 Nonlinear equilibrium strategy

Proposition 3. Suppose regulated buyers are asymmetric and risk-neutral. If they adopt the nonlinear quotation strategies, the nonlinear equilibrium quotation strategy for *i* is

$$d_i(p) = \overline{Q}(1 - \frac{p}{a'})^{\frac{1}{m-1}} - \frac{a'(\overline{Q} - \overline{q_i})}{a' - p} \tag{21}$$

where $\overline{q_i}$ is the initial demand for the regulated bidder *i* when $p = 0$, and there is no excess supply of quotas in the auction market due to $\overline{q_i} \ge \frac{uQ_0}{m}$.

Proof. Firstly, according to the first-order premise of maximizing the buyer's profit Eq.(7), we can get

$$\frac{d\left[\sum_{\substack{j=1\\j\ne i}}^{m} q_j\right]}{dp}(a'-p) + (uQ_0 - \sum_{\substack{j=1\\j\ne i}}^{m} q_j) = 0 \tag{22}$$

The simplified results are as follows:

$$\frac{d\left[\sum_{\substack{j=1\\j\ne i}}^{m} q_j\right]}{dp} = \frac{-q_i}{a'-p} \tag{23}$$

The sum of $i \in \{1, 2....m\}$ on both sides of equation (22), there is

$$(m-1)\frac{d\left[\sum_{i=1}^{m} q_i\right]}{dp} = \frac{-\sum_{i=1}^{m} q_i}{a'-p} \tag{24}$$

The solution of its ordinary differential equation can be obtained.

$$\sum_{i=1}^{m} q_i = A(a'-p)^{\frac{1}{m-1}} \tag{25}$$

Since when $p = 0$,the initial demand $\overline{Q}$, we get $A = \overline{Q}a'^{-\frac{1}{m-1}}$

$$\sum_{i=1}^{m} q_i = \overline{Q}(1-\frac{p}{a'})^{\frac{1}{m-1}} \tag{26}$$

where $q_i(p)$ can be referred to

$$q_i(p)=\sum_{i=1}^{m} q_i - \sum_{\substack{j=1\\ j\neq i}}^{m} q_j \tag{27}$$

Deriving on both sides of equation (26), we can get

$$\frac{dq_i}{dp} = [\overline{Q}(1-\frac{p}{a'})^{\frac{1}{m-1}}]' - \frac{d\left[\sum_{\substack{j=1\\ j\neq i}}^{m} q_j\right]}{dp} \tag{28}$$

Next, Substituting (22) into (27)

$$\frac{dq_i}{dp} = -\frac{\overline{Q}}{a'(m-1)}(1-\frac{p}{a'})^{\frac{2-m}{m-1}} - \frac{q_i}{a'-p} \tag{29}$$

Solving its ordinary differential equations, we obtain

$$q_i(p)=\overline{Q}(1-\frac{p}{a'})^{\frac{1}{m-1}} - \frac{w}{a'-p} \tag{30}$$

Substituting $q_i(0)=\overline{q_i}$ into (29), $w=(\overline{Q}-\overline{q_i})a'$ is obtained

$$d_i(p)=\overline{Q}(1-\frac{p}{a'})^{\frac{1}{m-1}} - \frac{a'(\overline{Q}-\overline{q_i})}{a'-p} \tag{31}$$

According to Eq.(28), $\frac{dq_i}{dp}<0$,so $d_i(p)$ is a monotonic decreasing function, and then

$$\frac{d^2R_i}{dp^2} = -\frac{d^2\left[\sum_{\substack{j=1\\ j\neq i}}^{m} q_j\right]}{dp^2}(a'-p^*)+2\frac{d\left[\sum_{\substack{j=1\\ j\neq i}}^{m} q_j\right]}{dp} \tag{32}$$

After deriving the Eq. (21) on both sides, there is

$$\frac{d^2\left[\sum_{\substack{j=1\\ j\neq i}}^{m} q_j\right]}{dp^2}(a'-p^*) = -\frac{dq_i}{dp} + \frac{d\left[\sum_{\substack{j=1\\ j\neq i}}^{m} q_j\right]}{dp} \tag{33}$$

Bring (31) into(30), and the result is as follow

$$\frac{d^2R_i}{dp^2} = \frac{d\left[\sum_{i=1}^{m} q_i\right]}{dp^*} \tag{34}$$

From the above $\frac{dq_i}{dp}<0$, we can see that $\frac{d^2R_i}{dp^2}\leq 0$, that is, the quotation strategy can satisfy the second-order condition of maximizing profit. Therefore, Eq.(20) is a nonlinear equilibrium quotation strategy.

Proposition 4. When the market reaches equilibrium (20), the government's optimal elastic allocation coefficient u^* and equilibrium price p^* for the regulated buyer are related to the initial demand ceiling $\overline{Q}$.

1) When $\overline{Q} \geq \frac{(Q_0 - Q_1)(a' - p_1)(a')^{m-1}}{m(a' - p_1)^{\frac{m}{m-1}} - (m-1)(a')^{\frac{m}{m-1}}}$, $u^* = \frac{\overline{Q}}{Q_0}[m(1 - \frac{p_1}{a'})^{\frac{1}{m-1}} - \frac{(m-1)}{a'(a' - p_1)}]$ $p^* = p_1$;

2) When $\overline{Q} < \frac{(Q_0 - Q_1)(a' - p_1)(a')^{m-1}}{m(a' - p_1)^{\frac{m}{m-1}} - (m-1)(a')^{\frac{m}{m-1}}}$, $u^* = \frac{Q_0 - Q_1}{Q_0}$, $p^*=p_2$, where p_2 is the positive solution of $[(1 - \frac{p_2}{a'})^{\frac{1}{m-1}} - \frac{(m-1)}{m(a' - p_2)a'}]=\frac{Q_0 - Q_1}{mQ_0}$.

Proof. According to the clearing condition at equilibrium $\sum_{i=1}^{m} d_i(p^*) = uQ_0$, where p^* is the equilibrium price, bring it into(29), and we get

$$uQ_0=m\overline{Q}(1 - \frac{p^*}{a'})^{\frac{1}{m-1}} - \frac{(m-1)\overline{Q}}{a'(a' - p^*)} \tag{35}$$

$$u = \frac{\overline{Q}}{Q_0}[m(1 - \frac{p^*}{a'})^{\frac{1}{m-1}} - \frac{(m-1)}{a'(a' - p^*)}] \tag{36}$$

It is easy to see from Eq. (34) that the elastic distribution coefficient u is a decreasing function concerning the equilibrium price p^* .

To ensure that non-regulated buyers participate in the auction market, the market equilibrium price shall not be higher than the highest quotation of the non-regulated buyers, that is, $p^*(u) \leq p_1$, and when the equilibrium price $p^*(u) = p_1$, we have

$$u^* = \frac{\overline{Q}}{Q_0}[m(1 - \frac{p_1}{a'})^{\frac{1}{m-1}} - \frac{(m-1)}{a'(a' - p_1)}] \tag{37}$$

The elastic distribution coefficient u for the regulated buyer needs to be met with $(1-u)Q_0 \leq Q_1$, then

$$u^* = \frac{\overline{Q}}{Q_0}[m(1 - \frac{p_1}{a'})^{\frac{1}{m-1}} - \frac{(m-1)}{a'(a' - p_1)}] \geq 1 - \frac{Q_1}{Q_0} \tag{38}$$

$\overline{Q}$ needs to satisfy

$$\overline{Q} \geq \frac{(Q_0 - Q_1)(a' - p_1)(a')^{m-1}}{m(a' - p_1)^{\frac{m}{m-1}} - (m-1)(a')^{\frac{m}{m-1}}} \tag{39}$$

Conversely, when $\overline{Q} < \frac{(Q_0 - Q_1)(a' - p_1)(a')^{m-1}}{m(a' - p_1)^{\frac{m}{m-1}} - (m-1)(a')^{\frac{m}{m-1}}}$, u does not satisfy the condition so that $p^*(u)$ reaches the maximum p_1 . In this case, $p^*(u) < p_1$, since the equilibrium price $p^*(u)$ is a monotone decreasing function of the elastic distribution coefficient u, and $p^*(u)$ takes the maximum at the minimum $u^* = \frac{Q_0 - Q_1}{Q_0}$, then $p^* = p_2$ is the positive solution of Eq.(37):

$$[(1 - \frac{p^*}{a'})^{\frac{1}{m-1}} - \frac{(m-1)}{m(a' - p^*)a'}]=\frac{Q_0 - Q_1}{mQ_0} \tag{40}$$

where p_2 is the positive solution of $[(1-\frac{p_2}{a'})^{\frac{1}{m-1}}-\frac{(m-1)}{m(a'-p_2)a'}]=\frac{Q_0-Q_1}{mQ_0}$, when m、a'、Q_1、Q_0 are known, they can be solved by computational software, like Maple.

Proposition 5. The proposed uniform-price auction mechanism of carbon emission rights based on flexible supply is a feasible auction mechanism, which meets the constraints of buyers' participation and incentive compatibility.

Proof. First, assuming that when the regulated buyer adopts a linear or nonlinear quotation strategy, the equilibrium price, the allocated amount, and the buyer's income can be expressed as $p^{(1)}$ and $p^{(2)}$, $q^{(1)}$, and $q^{(2)}$, respectively. Since $p^{(1)}\le p_1<a', p^{(2)}\le p_1<a', q^{(1)}\ge 0, q^{(2)}\ge 0$, we can easily conclude that $R_i^{(1)}>0, R_i^{(2)}>0$ according to the profit function $R_i=z_i+(a'-p)q_i$. So, for any bidding price *p*, we have $R_i\ge 0$, indicating that this mechanism satisfies the buyer's rationality and participation constraints.

Secondly, according to the above discussion, the equilibrium price p^* is related to the upper limit of the initial demand $\overline{Q}$. Under the linear and nonlinear equilibrium quotation strategy, if $\overline{Q}\ge(Q_0-Q_1)(\frac{1}{a'-p}+m-2)$ and $\overline{q}_i\ge\frac{(Q_0-Q_1)}{m}(\frac{1}{a'-p}+m-2)$ or $\overline{Q}\ge\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m(a'-p_1)^{\frac{m}{m-1}}-(m-1)(a')^{\frac{m}{m-1}}}$ and $\overline{q}_i\ge\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m^2(a'-p_1)^{\frac{m}{m-1}}-m(m-1)(a')^{\frac{m}{m-1}}}$ are satisfied, the ideal equilibrium solution (the government gains more)can be achieved. For the regulated companies, the profit of regulated buyer *i* is $R_i=z_i+a'q_i-pq_i$. It is easy to find that income is positively related to the number of declarations. The larger the declaration amount, the more the benefits. Therefore, in the proposed auction model, regulated companies have the incentive to report their own demand truthfully, and the constraint of incentive compatibility is established. In summary, the auction method is described to satisfy both the participation and incentive compatibility constraints. It can avoid the problem of "demand hiding" in the regulated buyers and is a feasible auction mechanism.

EXTENDED MODEL ANALYSIS

Corollary 1. When the regulated buyers adopt the linear quotation strategies, there are multiple equilibrium points. When $p_1\le a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}$, the equilibrium point is better for the government.

Proof. According to Proposition 2, when $p_1\le a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}$, $u^*=\frac{(a'-p_1)(m-1)}{(a'-p_1)(m-2)+1}\times\frac{\overline{Q}}{Q_0}$, and $p^*=\min\{p_1, a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}\}=p_1$; on the contrary, $u^*=\frac{Q_0-Q_1}{Q_0}$, $p^*=a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}$ there are two equilibrium points. In these two cases, the government revenue is $\Pi^{(1)}$ 和 $\Pi^{(2)}$. Compare two equilibrium prices, the former $p^*=p_1$, and the latter one $p^*=a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}<p_1$, $\Pi^{(1)}>\Pi^{(2)}$. Therefore, for the government, when $p_1\le a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}$ is satisfied, the government's income is higher, and the clearing price is closer to the real marginal abatement cost of the regulated buyers. This equilibrium point is better for the government.

Corollary 2. The auction mechanism proposed in this paper can overcome the low price equilibrium between $[0,\min\{p_1,a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}\}]$.

Proof. As can be seen from the proof of Corollary 1, the equilibrium price p^* changes due to the change in the highest bidding price of the unregulated buyer p_1. Therefore, the equilibrium price generated by this auction mechanism will not be lower than $\min\{p_1,a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}\}$, and the equilibrium price must be non-negative, which can prove the conclusion of Corollary 2.

Corollary 3 . When the regulated buyers adopt nonlinear quotation strategies, there are multiple equilibrium points when $\overline{Q}\ge\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m(a'-p_1)^{\frac{m}{m-1}}-(m-1)(a')^{\frac{m}{m-1}}}$, the equilibrium is better for the government.

Proof. According to Proposition 4, the equilibrium price p^* is related to the initial demand ceiling $\overline{Q}$.

When $\overline{Q}\ge\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m(a'-p_1)^{\frac{m}{m-1}}-(m-1)(a')^{\frac{m}{m-1}}}$, is satisfied, equilibrium price $p^*=p_1$; otherwise, when $\overline{Q}<\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m(a'-p_1)^{\frac{m}{m-1}}-(m-1)(a')^{\frac{m}{m-1}}}$, elastic distribution coefficient u takes the minimum $u^*=\frac{Q_0-Q_1}{Q_0}$, and $p^*<p_1$. It can be seen that the equilibrium point of the nonlinear equilibrium strategy is not unique. In two cases, the government revenue is $\Pi^{(1)}$ and $\Pi^{(2)}$ respectively. Comparing the two equilibrium prices, the former $p^*=p_1$ and the latter $p^*<p_1$, it is easy to get $\Pi^{(1)}>\Pi^{(2)}$, so for the government, when $\overline{Q}\ge\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m(a'-p_1)^{\frac{m}{m-1}}-(m-1)(a')^{\frac{m}{m-1}}}$, the government revenue is larger, and the clearing price is closer to the real marginal emission reduction cost of the emission regulated buyers. Therefore the equilibrium achieved at this time is better.

Corollary 4. The above auction mechanism can overcome the low-price equilibrium between $[0,\min\{p_1,p_2\}]$, where p_2 is the positive solution of the equation $[(1-\frac{p_2}{a'})^{\frac{1}{m-1}}-\frac{(m-1)}{m(a'-p_2)a'}]=\frac{Q_0-Q_1}{mQ_0}$.

Proof. From the proof of corollary 3, we can see that the equilibrium price p^* changes when $\overline{Q}$ is different, and it is easy to conclude that the equilibrium price in the proposed auction will not be lower than $\min\{p_1,p_2\}$, where p_2 is the positive solution of the equation

$[(1-\frac{p_2}{a'})^{\frac{1}{m-1}}-\frac{(m-1)}{m(a'-p_2)a'}]=\frac{Q_0-Q_1}{mQ_0}$; meanwhile, equilibrium price p^* must be non-negative, which proves corollary 4.

Considering that the implicit collusion and low-price equilibrium caused by potential "buyer threat" under fixed supply, this paper proposes a uniform-price auction mechanism of carbon emission right based on elastic supply under the asymmetric scenario, using elastic allocation coefficient to optimize the total allocation of carbon emission right of regulated buyers. Based on the above conclusions, this paper further derives valuable conclusions for government decision-making. Then, from the perspective of maximizing government revenue and stimulating the quotation strategy of regulated buyers to report closer to the real marginal emission reduction cost, we put forward some suggestions on the design of the government auction mechanism.

Corollary 5. The quotation strategy of the regulated buyer is a convex non-increasing function, which is a prerequisite for the existence of a linear or nonlinear equilibrium quotation strategy.

Proof. Propositions 1 and 3 show that only when $\frac{dR_i^2}{dp^2}<0$ and $\frac{dq_i}{dp}<0$, there can be a second-order condition of maximum profit when regulated bidder i adopts a strategy to deal with other players. That is, when the bidding strategy of other buyers is known, regulated buyer i will not deviate from the equilibrium. At this time, the optimal price is the equilibrium price $p(u)^*=\arg\max_p\{R_i=z_i+(a'-p)(uQ_0-\sum_{\substack{j=1\\j\neq i}}^{m}q_j)\}$.

Corollary 6. Whether the regulated buyer adopts a linear or nonlinear equilibrium quotation strategy, the equilibrium price p^* is negatively correlated with the elastic allocation coefficient u , and the smaller the proportion of carbon allowances allocated to the regulated buyers, the higher the equilibrium price.

Proof. It is easy to derive $\frac{d[p^*(u)]}{du}<0$ under linear and nonlinear equilibrium quotations from Eq.(17) and (34), so the equilibrium price p^* is a decreasing function of the elastic partition coefficient u . This is consistent with our common sense. The smaller the total amount allocated to the regulated companies, the greater the competition between them, encouraging enterprises to quote more closely to their real marginal emission reduction costs.

Corollary 7. Whether the regulated buyer adopts the linear quotation strategy (3) or the nonlinear quotation strategy (20), if the number of non-regulated buyers m is large enough, there is a unique equilibrium $p^*=p_1$ in the auction model.

Proof. First, it can be known from Proposition 5.2 that when the regulated buyer adopts the linear equilibrium strategy (5-3), $p^*=\min\{p_1,a'-\frac{Q_0-Q_1}{(m-1)\bar{Q}-(m-2)(Q_0-Q_1)}\}$ if the number of non-regulated buyers m is large enough, $\frac{Q_0-Q_1}{(m-1)\bar{Q}-(m-2)(Q_0-Q_1)}\to 0$, and as $p_1\le a'$, we have $p^*=\min\{p_1,a'-\frac{Q_0-Q_1}{(m-1)\bar{Q}-(m-2)(Q_0-Q_1)}\}=p_1$; for the same reason, if $Q_1\to Q_0$, $\frac{Q_0-Q_1}{(m-1)\bar{Q}-(m-2)(Q_0-Q_1)}\to 0$, $p^*=\min\{p_1,a'-\frac{Q_0-Q_1}{(m-1)\bar{Q}-(m-2)(Q_0-Q_1)}\}=p_1$.

Secondly, it can be known from Corollary4 that when regulated buyers adopt the nonlinear quotation strategy (20), and if m approaches infinity or the positive solution p^* of $[(1-\frac{p^*}{a'})^{\frac{1}{m-1}}-\frac{(m-1)}{m(a'-p^*)a'}]=\frac{Q_0-Q_1}{mQ_0}$ is approaching $a'-\frac{1}{a'}$, we obtain $p^*=\min\{p_1,a'-\frac{1}{a'}\}=p_1$, thus auction model only has a unique equilibrium, namely $p^*=p_1$.

Corollary 8. Whether the buyer adopts a linear equilibrium quotation strategy (3) or a nonlinear equilibrium quotation strategy (20), as long as $\overline{q_i}\geq\max\{\frac{uQ_0}{m},\frac{(Q_0-Q_1)}{m}(\frac{1}{a'-p}+m-2),\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m^2(a'-p_1)^{\frac{m}{m-1}}-m(m-1)(a')^{\frac{m}{m-1}}}\}$, the market always achieves the ideal equilibrium solution.

Proof. According to Proposition 1, we get $\overline{q_i}\geq\frac{uQ_0}{m}$, and Corollary 1 tells us, under the linear equilibrium quotation strategy, when $p_1\leq a'-\frac{Q_0-Q_1}{(m-1)\overline{Q}-(m-2)(Q_0-Q_1)}$ is satisfied, we have $p^*(u)=p_1$, and $\overline{Q}\geq(Q_0-Q_1)(\frac{1}{a'-p}+m-2)$ can be calculated, and since $\overline{Q}=\sum_{i=1}^{m}\overline{q_i}$, so when $\overline{q_i}\geq\frac{(Q_0-Q_1)}{m}(\frac{1}{a'-p}+m-2)$ is satisfied, the market reaches the optimal equilibrium price $p^*=p_1$; in addition, if the nonlinear equilibrium quotation strategy is adopted, when $\overline{Q}\geq\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m(a'-p_1)^{\frac{m}{m-1}}-(m-1)(a')^{\frac{m}{m-1}}}$ is true, the equilibrium price takes the maximum value $p^*=p_1$, at this time $\overline{q_i}$ need to meet the condition $\overline{q_i}\geq\frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m^2(a'-p_1)^{\frac{m}{m-1}}-m(m-1)(a')^{\frac{m}{m-1}}}$. Above all, if $\overline{q_i}$ meets with the above three conditions simultaneously, the market can achieve the optimal equilibrium price under the uniform-price auction of carbon allowance, no matter under the linear equilibrium strategy or the nonlinear equilibrium strategy.

According to corollary 5~8, this paper proposes four suggestions based on the uniform-price auction mechanism of carbon emission rights with flexible supply, to guide the market to achieve an ideal equilibrium and maximize government revenue.

(1) Reduce the entry threshold of regulated buyers and encourage more regulated enterprises to enter the auction market.

If the number of regulated buyers is large enough, the auction model has a unique equilibrium, at which time the market achieves an ideal clearing price $p^*=p_1$. (Corollary 7)

(2) Reduce the participation restrictions on non-regulated buyers (institutional investors and individual investors), and increase the total carbon allowance of non-regulated buyers flexibly.

The equilibrium price p^* is negatively correlated with the elastic allocation coefficient u. The smaller the proportion of allocation to regulated buyers, the fiercer the competition between them, which also encourages the price quotation of emission control enterprises to be closer to their true marginal emission reduction costs. (Corollary 6)

(3) Design of auction rules: The regulated buyer's declaration function must be a convex function to avoid a low-price worse equilibrium. Only when the buyer's quotation strategy is a convex

function, the equilibrium quotation strategy exists in the auction market. Under this strategy, low-cost equilibrium can be avoided. (Corollary 5)

(4) Initial declaration limit: guide the buyer's initial auction volume $\overline{q_i}$ to meet with $\overline{q_i} \geq \max\{\frac{uQ_0}{m}, \frac{(Q_0-Q_1)}{m}(\frac{1}{a'-p}+m-2), \frac{(Q_0-Q_1)(a'-p_1)(a')^{m-1}}{m^2(a'-p_1)^{\frac{m}{m-1}}-m(m-1)(a')^{\frac{m}{m-1}}}\}$.

When the amount of declaration $\overline{q_i}$ meets the above conditions, the optimal equilibrium price can be achieved under either linear or nonlinear equilibrium bidding strategy $p^*=p_1$, and the government's revenue is the largest at this time. (Corollary 8)

CONCLUSIONS

To overcome the problems of implicit collusion and low-price equilibrium, this study improves the current uniform-price auction mechanism of carbon emission rights based on a fixed amount. It proposes a uniform price auction mechanism with a flexible supply. Differing from the traditional single type of buyers, we introduce non-regulated buyers into the uniform-price auction of carbon emission rights to promote market activity and effective emissions reduction. The auctioneer determines the elastic supply quantity $uQ_0(0<u<1)$ after collecting the bidding demand of the regulated buyers. In the construction of the model, we break through the assumption of "common valuation, same distribution" of the buyers in the traditional auction mechanism studies and expands into a more practical situation, where the regulated and non-regulated buyers are independent of each other and have a private valuation. Therefore the applicable scenario is more extensive. Besides, this paper solves both linear and nonlinear equilibrium in the case of asymmetric regulated buyers and further analyses equilibrium price and elastic distribution coefficient under different equilibrium strategies. The results show that there are multiple equilibriums in the auction mechanism. The auction is guided to an ideal equilibrium point (the government earns the most) to avoid the negative effects of multiple equilibrium points of the seller's earnings. We put forward specific suggestions from four aspects: market access threshold, initial carbon allowance declaration volume, and buyers' concave and convex bidding strategies. Finally, it is proved that the auction mechanism satisfies the constraint conditions of participation and incentive compatibility. Therefore, the proposed auction is an effective auction mechanism, and it can effectively suppress the occurrence of implicit collusion and low-price equilibrium.

REFERENCES

Anderson, B., & Maria, C. D (2011). Abatement and allocation in the pilot phase of the EUETS. Environmental & Resource Economics, 48(1): 83-103.

Ausubel L M, Cramton P, Pycia M (1998).Demand Reduction and Inefficiency in Multi-Unit Auctions. The Review of Economic Studies.81 (4): 1366–1400.

Cramton P. (1998). Ascending auctions. European Economic Review, 42 (3):745-756.

Chesney, Marc; Taschini, Luca; Wang, Mei(2011). Regulated and non-regulated companies, technology adoption in experimental markets for emission permits, and option contracts. Grantham Research Institute on Climate Change and the Environment 41, University of Zurich.

Cui, L. B., Fan, Y., Zhu, L., & Bi, Q. H (2014). How will the emissions trading scheme save cost for achieving china's 2020 carbon intensity reduction target? Applied Energy, 136(12):1043-1052.

Fan X. Y (2005). Multi-item auction under variable supply. Operations Research and Management. 14(2):70-74.

Gretschko V, Rajko A (2015).Excess information acquisition in auctions. Experimental Economics.18 (3):335-355.
Hitzemann S, Uhrig-Homburg M, Ehrhart K M (2015). Emission permits and the announcement of realized emissions: Price impact, trading volume, and volatilities. Energy Economics, 51(5):560-56.
Hu, D.B., Hu, Z.Y., Chen, X.H (2019). Influence of Agent-based bidding learning on carbon emission rights auction. Journal of Systems Engineering. 34 (02):170-185.
Klemperer P D, Meyer M A (1989). Supply Function Equilibria in Oligopoly under Uncertainty. Econometrica. 57(6):1243-1277.
Liu W. Y (2015). Research on the Countermeasures of Carbon Pilot Operation in Seven Provinces and Cities of China. Economic Research Guide. 240(24):176-177.
Li, Y (2015). The valuation relevance of greenhouse gas emissions under the European Union carbon emissions trading scheme. European Accounting Review, 24(3): 551-580.
Min X J, Dong X Y, Zi Y C (2016).Market power in auction and efficiency in emission permits allocation. Journal of Environmental Management. 183(3):576-584.
Newell, R. G., Pizer, W. A., & Raimi, D (2013). Carbon markets 15 years after Kyoto: lessons learned, new challenges. Journal of Economic Perspectives, 27(1): 123-146.
Parry, I. W. H., Iii, R. C. W., & Goulder, L. H (1999). When can carbon abatement policies increase welfare? The fundamental role of distorted factor markets.Journal of Environmental Economics & Management, 37(1):52-84.
Rao C.J., Zhao Y., Wang Q (2012). Separation of Separable Goods Based on Variable Supply and Its Application. Chinese Management Science, 20(1): 129-138.
Rao C.J., Zhao Y, Li C (2012). Asymmetric Nash equilibrium in emission rights auctions. Technological Forecasting and Social Change, 2012, 79(3):429-43.
Taschini L, Chesney M, Wang M (2014). Experimental comparison between markets on dynamic permit trading and investment in irreversible abatement with and without non-regulated companies. Journal of Regulatory Economics, 46(1):23-50.
Wang J J D, Zender J F (2002). Auctioning divisible goods. Economic Theory, 19(4):673-705.
Wilson R (1979). Auctions of Shares. Quarterly Journal of Economics. 93(4):675-689
Wang S.F., Yang S.L (2016). Carbon Emission Rights Auction Considering the Influence of Reserve Price on Bidding Strategies. Journal of Management Engineering. 30(2):181-187.
Wang Y, Bi Z.W, Li C.L (2004). Effect of commission on auctions. Journal of Management Science, 7(4):45-48.
Zhang Y J, Wang A D (2015). Tan W. The impact of China's carbon allowance allocation rules on the product prices and emission reduction behaviors of ETS-covered enterprises. Energy Policy. 86(1):176-185.

DECISION SCIENCES INSTITUTE

Cognitive Processes of the Boundedly Rational Newsvendor

Tarikere T. Niranjan
Indian Institute of Technology Bombay
Email: ttniranjan@iitb.ac.in

Narendra K. Ghosalya
Indian Institute of Technology Bombay
Email: narendraghosalya@gmail.com

Kristian Rotaru
Monash University
Email: kristian.rotaru@monash.edu

Srinagesh Gavirneni
Cornell University
Email: nagesh@cornell.edu

Jayanth Jayaram
University of South Carolina
Email: jayaram@moore.sc.edu

ABSTRACT

Controlled experiments using eye tracking uncover novel nuances of psychophysiological processes behind the boundedly rational newsvendor. Interventions include ordering multiple products once instead of the same product at multiple instances, standing orders, sharing the critical fractile logic, training participants (students and managers) on the newsvendor logic, decomposed newsvendor task and the interplay of in-stock probability and fill rate.

KEYWORDS: Decision making under uncertainty, Eye tracking, Controlled experiments, Forecasting

DECISION SCIENCES INSTITUTE
Managing Risks in Circular Economy Using Flexibility Based Strategies

Divya Choudhary
Indian Institute of Management Sambalpur, Odisha, India
divyachoudhary2626@gmail.com

"ABSTRACT"

Increasing number of organizations are incorporating circular economy (CE) practices in their business to sustain in the competitive market. CE is multifaceted and associated with various risks. The study aims to explore the role flexibility based strategies in managing risks associated with CE practices. The literature addressing the role of flexibility in CE risk mitigation is very limited. Accordingly, a CE risk management framework based on flexibility is proposed to handle the CE risks. The framework presents the effects of various dimensions of flexibility on the identified CE risks and sets the significant foundation for future research in this direction.

KEYWORDS: Circular Economy (CE), Flexibility, Framework, Risk Mitigation, Strategies

INTRODUCTION

Organizations are progressing towards implementing sustainable practices as a result of legislative pressure, consumer awareness and increasing green issues (Krike et al. 2013; Singh & Agrawal, 2018). Value reclamation of end-of-life, end-of-use, obsolete and warranty returns is one such practice, which constitutes the Circular Economy (CE) (Ullah & Sarkar, 2020). CE comprises backflow of products from consumers to manufacturers in order to capture the value from returns through recovery operations such as reuse, recycle, resale, remanufacture etc. The scope of return system is greater than just addressing the legal and environmental responsibilities, as it also provides financial benefits and expands the competitive abilities of the organization (Choudhary & Madaan, 2013). It also assists in the structuring of recovery strategies that can enhance organization performance and increase product sales (Prahinski & Kocabasoglu, 2006). Moreover, the amount of products entering the reverse channel is rising rapidly, as returns worth of about $100 billion value are being returned every year (Stock, 2001). Even in a developing country such as India, e-waste generated is increasing with a growth rate of 10 percent every year, which poses huge risks to the community due to the presence of hazardous elements (Agrawal et al. 2014). Accordingly, it is very essential for the organizations to take care of the returns and the wastes induced. A body of knowledge is beginning to develop around the product returns field, which has emerged within the last one decade or two. Especially during the last decade, CE has obtained recognition both as a research field and as a practice (Rogers & Tibben-Lembke, 2001; Farhani et al. 2019).

CE can be defined as an extension of conventional supply chain to handle returns for optimum utilization of natural resources and to impede environmental degradation (Ullah & Sarkar, 2020). In other words, the set of processes which can incorporate the flow of returns beyond the conventional supply chain scheme is known as CE. According to Guide & Van Wassenhove (2002), CE encompasses five operations: Product Acquisition (Collection of returns from customers), Reverse Logistics (Gate-keeping, Transportation, Warehousing, and Inventory management), Inspection and disposition (Assessment of returns for suitable recovery option selection), Reconditioning (Value reclamation from returns through operations such as repair, cannibalization, remanufacturing and recycling etc.) and Distribution and sales (Developing secondary markets for reclaimed products).

The flow of returns has very less visibility compared to the conventional supply chains, which results in surfacing of large number of uncertainties in CE (Bai & Sarkis, 2013). The uncertainties are related to quality, time, amount and diversity of returns; estimation of operation and cost related parameters for reverse logistics networks; decisions about resolution for product returns and costs of co-ordination along the return system etc. (Srivastava, 2008; Fleischmann, 1997; Rogers & Tibbn-Lembke, 1999). These uncertainties make the CE vulnerable and can trigger an unplanned event, which can have unwanted consequences. Accordingly, it can be deduced that product recovery is a multifaceted process and along with many advantages, also brings along a number of risks in the business practices (Choudhary et al. 2020). It is essential to mitigate these risks associated with CE for enhancing its efficiency and achieving the desired results (Janse et al. 2010; Farhani et al. 2019).

To help manage these risks and uncertainties, it is required to build flexibility in CE system (Bai & Sarkis, 2013). Flexibility deployment in CE can handle the uncertainties and possibilities of disruptions in a reverse channel (Tang & Tomlin, 2008). However, considerable amount of research has been done the area of risk management and flexibility in conventional supply chains; these areas are relatively unexplored and immature in the domain of CE. Moreover, research on role of flexibility in CE risk management has virtually seen no study compared to other aspects. The study attempts to address this gap in literature building upon the research in conventional supply chain. The paper initially identifies the risks existing in return system and then proposes a flexibility framework presenting the role of various dimensions of flexibility in mitigating the same. The framework is presented in tabular form and builds upon the study of risk management and flexibility in conventional supply chain. The research has practical implications for the organizations involved in product recovery and/or planning to incorporate flexibility in CE. It can assist the organizations in controlling the recovery channel risks and will guide the managers to incorporate flexibility in CE. The contribution of the paper is threefold: 1) Identification of risks in product returns system; 2) Introducing a flexibility framework to mitigate CE risks; 3) Discovery of future research stream of flexibility and risk management in CE. In the next section, CE literature is discussed and various associated risks are identified. The risk mitigation framework is proposed in the following section.

LITERATURE REVIEW

Rogers & Tibben-Lembke (1999) suggested that organizations can increase their efficiency through systematic incorporation of recovery operations in their existing systems. Accordingly, as environmental awareness is increasing, organizations are accommodating value reclamation operations, which constitute the Circular Economy (CE) in order to gain competitive advantage (Srivastava & Srivastava, 2006). CE is a value recovering system in which value of end-of life, end-of-use, defected, obsolete and warranty reruns is retrieved with the help of various product recovery processes (Wadhwa et al. 2009). The scope of CE has evolved over time for several reasons as follows: increase in amount of returns up to 50% of sales; enforcement of legislative regulations; consumer awareness; and rising cost of landfill facilities (Prahinski & Kocabasoglu, 2006; Singh & Agrawal, 2018). Additionally, efficient management of returns results in value addition, conservation of virgin resources and sustainable development of the organizations in a sound manner Moreover, recovery of returns has become a primary managerial focus from a cost view point and the impending influence on consumer loyalty (Madaan et al. 2012). In a survey, it was found that 82 percent of customers are not willing to shop again in case of inconvenient return policies (Skinner et al. 2008). Accordingly, organizations are required to strategically manage CE to enhance their performance in today's highly competitive market.

Despite the advantages, management of CE is complicated due to involvement of multifunctional contextual environment of many return processes (Choudhary et al. 2014). It has been estimated that a returned product looses about 45% of its strategic as it moves through the reverse channel (Blackburn et al. 2004).This loss can be even higher for the short life cycle products such as laptops etc., which can lose value in excess of 1% per week. One of the major reasons for this tremendous loss of asset is the product's decrease in value with time (Gobbi, 2011). This elapsed time in value recovery is the result of various reasons such as: preoccupation with core business; Lack of awareness; intricacies in End-of-life decision making; ambiguities and difficulties in collection and transportation of returns; uncertainties associated with the cost involved and revenue (Blackburn et al. 2004); and changing government regulations. Uncertainty of supply and timings further enhances the complexity of the system. Return channels for most of the products are relatively immature and uncertain, which exposes CE to large number of risks that disrupts the flow of return. Accordingly, it is required to focus on mitigation of uncertainties and risks present in CE for reducing the negative implications of the occurrence of certain associated disruptions.

A substantial amount of previous research has focused on various CE issues such as environmental aspects (Stock 2001); network design (Srivastava, 2008); third party reverse logistics provider (Meade & Sarkis, 2002); profitability (Tan & Kumar 2006); inventory models (Jaber & El Sadany, 2009); optimal EOL alternative selection (Wadhwa et al. 2009); and model for CE adopters (Barker & Zabinsky, 2011) etc. However, literature explicitly focusing on the issue of risk control in CE is virtually nonexistent. To fill this gap in literature, we introduce a flexibility based framework to curb the impacts of risks existing in value chain of returns. Initially, it is required to discover the risks that posses threat to CE.

According to the Association of Project Managers, risk is defined as an uncertain event or set of circumstances which can have an effect on achievement of one or more objectives (Tand & Tomlin, 2008). Nowadays, organizations have become a complicated and sophisticated system, which consists of various operations that are exposed to risks and vulnerabilities (Choudhary et al. 2020). Controlling these risks is necessary for the organizations to accomplish their goals and for enhancing their market performance. Strategic-minded organizations have acknowledged that risks can never be completely eliminated; however their realization and consequences can be minimized (Choudhary et al. 2021). Hence, the traditional perspective of avoiding risks is changing and organizations are attempting to manage their risk exposures so that, they incur just the adequate amount of the right kind of risks, which do not hamper the performance and progress. For the above mentioned reasons, it is essential for the organizations to control the risks in a way towards the achievement of overall goals. To accomplish this, a risk management process is required that is practical, sustainable, and easy to understand.

The process of risk management usually begins with a study to identify the potential risks that a system could generate or to which it could be exposed (Gandhi et.al, 2012). Further, as CE is an extended part of conventional supply chain, there are chances to overlap of risks among the two. Also, an improvement in CE performance can enhance the efficiency of forward value chain. Accordingly, structuring upon the literature of supply chain risks and CE, some of the CE risks have been identified as presented in Table 1.

Table 1: Risks in Circular Economy

Risk Categories	Risk Events	Description
Returns Collection Risk	Collection point risk	Risk associated with the location of collection points i.e. weather they are accessible to all range of customers willing to return products.
	Gate keeping Risk	Risk associated with the erroneous screening of return products and allowing the undesired returns to enter the product return system.
	Product Returns Forecast Risk	Risk due the inability to predict the time and amount of return products.
	Transportation Risk	Risks associated with the transportation of the returned goods, such as delays due to bad weather, traffic density and breakdown, damage of return products during transportation, and accidents.
	Return Handling Risk	Risk related to the safety of workers while handling toxic returns and chances of damage during inventory storage in warehouses.
Inspection and Disposition Risks	Information Risk	Uncertainty regarding access to accurate information about the return products such as constituents, harmful effects if any, list of component parts etc. required for their reprocessing.
	Marginal Value of Time Risk	Risk due to the inability to process the returned products within the desired time which can in turn lead to deterioration of residual value of returns.
	Decision Making Risk	Risk associated with the selection of appropriate processing technique such as recycling, reuse, cannibalization, disposal etc. for recapturing the value of return products.
	Facility Risk	Risk due to the breakdown of machine, electricity or water failure causing a delay in recovery operations or leading to unavailability of plants, warehouses and official buildings.
Reprocessing Risks	Lower Recovery Rate Risk	Risk of lower-than-expected recovery from the return products which can affect the plant's effective production capacity.

	Quality of Reprocessed Output Risk	Quality related issues of remanufactured, repaired, secondary products and products manufactured from recycled raw material.
	Disassembly Complicatedness Risk	Uncertainty about the disassembly sequence of the return leading to wastage of time and sometimes inability to disable the product due to the complexity involved.
	Sustainable Regulations and Compliance Risk	Risk related to consequences of not complying with the appropriate regulations (local and global) and uncertainty about changing return policies.

The various acknowledged risk issues are the cause of disruptions in the recovery channel and to control them, it is required to implement flexibility in the CE and its operations (Tang & Tomlin, 2008). Flexibility has been evolved in literature as a coping mechanism to empower uncertainty in an organization (Sawhney, 2006). The role of flexibility in mitigating CE risks and the proposed framework are discussed in the next section.

FLEXIBILITY BASED FRAMEWORK FOR CE RISK MITIGATION

Flexibility can be defined as an adaptive response to sudden environmental variances (Purvis et al. 2014). In other words, it is capability of a system to transform its states or react quickly with least efforts and cost and without hampering the performance. Each flexibility domain is comprised of three elements: range, response and uniformity (Koste et al. 2004). The range element refers to number of different states and situation a system can adapt to. The ease and efficiency with which a system changes from one state to another is represented by response element. Any alteration or deterioration experienced by a system during achievement of one state from another falls under the uniformity element of flexibility. Flexibility develops resilience in an organization, which enables it to tackle risks swiftly and efficiently (Tang & Tomlin, 2008). The level of flexibility required depends upon the uncertainty involved in the industry, operations, product, market etc under consideration (Madaan & Wadhwa, 2007).

The return channel has a highly uncertain environment, which exposes it to a number of risks as identified in the previous section. These disruptions can be due to various reasons but, in any case return system should be flexible enough to recover at the earliest and with least efforts. Flexibility enables a system to respond quickly and efficiently to changing environments. Consequently, Flexibility deployment is essential in CE to mitigate the risks present due to unpredictable, vulnerable and uncertain flow of returns (Bai & Sarkis, 2013). CE flexibility refers to promptness of recovery system to overcome risk and its ability to adjust recovery speed, volume capacity and heterogeneity in response to changes in return channel. Apart from controlling the recovery system risks and uncertainties, flexibility can also be used to enhance the versatility and performance of the CE (Wadhwa et al. 2008).

To illustrate the importance of flexibility in risk mitigation, a framework has been introduced signifying the role of various dimensions of flexibility in controlling CE risks. In the literature, studies have focused on different dimensions of flexibility such as new product, labor, volume, delivery, access, mix, material handling, machine and process etc. (Tang & Tomlin, 2008; Sawhney 2006; Wadhwa et al. 2008; Koste et al. 2004; Purvis et al. 2014). We have focused on five major categories of flexibility, which are perceived as the most relevant by prior researchers;

Product mix (ability to recover value from different range of returns), Expansion (range and ability to accommodate changes in handling capacity of returns), Equipment (ability to perform wide range of reprocessing operations on different kinds of returns), Scheduling (ability to reschedule recovery operations) and Labor (ability of workers to handle a range of tasks and to adapt situational changes). Other dimensions of flexibility can also be considered depending upon the returns and industry. The framework is presented in a tabular form with proper categorization to show how each flexibility category is assisting the CE to respond to various identified risks. Further, the operational aspects of the framework are build upon the return channel and flexibility literature, whereas, the strategic aspects have been derived from the supply chain risk management literature. The flexibility based framework for CE risk mitigation is shown below in Table 2.

Table 2: Flexibility Based Framework for CE Risk Mitigation

Risks		Dimensions of Flexibility				
Risk Categories	Risk Events	Product Mix	Expansion	Equipment	Scheduling	Labor
Returns Collection Risk	Collection point risk	Ability to collect all ranges of returns	Capacity to collect various amount of returns	Equipment capability to handle different types of product returns	Ease of changing collection timings according to requirement	Versatility to handle all type customer and returns
	Gate keeping Risk	Able to sort different ranges of returns and materials with various specifications	Capability to sort small and large size batches of returns efficiently.	Ability to separate same type of returns in different ways.	Flexibility to reschedule and change the pace of sorting.	Labor flexibility to accommodate all changes
	Product Returns Forecast Risk		Capacity to handle and store different lot sizes of returns economically	Machine flexibility to process various lot sizes of returns effectively without delays.	Ability to schedule processing of returns in accordance with amount and time of their arrival.	Labor flexibility to accommodate all changes
	Transportation Risk	Transportation flexibility to carry returns of different sizes, delicacies and ranges safely.	Ability to transport different batch sizes of product returns economically.		Flexibility of transportation vehicles to change their routes according to the situation	Labor flexibility to accommodate all changes
	Return Handling Risk	Capability to manage returns with various material compositions		Flexibility to perform a wide range of operations to recover returns with	Ability to change the processing route and schedule of returns	Capability of workers to deal with returns having wide range of

		safely		different material compositions efficiently	according to their composition	compositions
Inspection and Disposition Risks	Information Risk	Ability to handle returns with different components and working conditions.		Enables equipments to handle returns with different parts and quality	Easy Changing of recovery route of returns according to their quality	Labor flexibility to handle returns with all kind of constituent parts and quality
	Marginal Value of Time Risk	Facilitate recovery of returns with different value erosion rates efficiently	Expansion flexibility to reclaim different volumes of returns in same time to avoid delays	Equipment ability to vary processing speed for different value erosion rates	Allows scheduling of returns according to their value erosion rates	Labor flexibility to accommodate all changes
	Decision Making Risk	Enables best End-of-Life option selection for different types of returns.	Ability to make proper recovery decisions for different volumes of returns		Changing the schedule of End-of-Life decision making as per required	Capability of workers to make correct decision for all the returns
Reprocessing Risks	Facility Risk		Able to manage small and large capacities of products economically even in case of breakdown	Flexibility to perform wide range of operations with varying speeds to make up for a machine failure	Flexibility to reschedule in case of machine breakdown	Labor flexibility to accommodate all changes
	Lower Recovery Rate Risk		Allows to deal with varying amounts of material recovery without hampering production		Facilitates changes in production schedule according to varying rates of recovery	Labor flexibility to accommodate all changes
	Quality of Reprocessed Output Risk	Flexibility to reprocess different range of input returns into high quality secondary products	Capacity to handle different volumes of returns without affecting the output quality	Ability to produce same quality secondary products from varying quality input returns		Labor capability to accommodate all changes in order to ensure high quality secondary

						products
	Disassembly Complicatedness Risk	Enables handling of simple to complicated assemblies	Flexibility to disassemble different lot sizes of returns with same efficiency	Facilitate disassembling of different products in different ways	Ability to change disassembly schedules	Labor flexibility to accommodate all changes
	Sustainable Regulations and Compliance Risk	Flexibility to comply with changing regulations associated with wide range of returns	Ability to comply regulations for all returns irrespective of their lot sizes (amount)			Flexibility of workers to incorporate changes according to changing regulations

The identified risks are present in various phases of CE and are essential to manage for the proper functioning of recovery systems in order to achieve the desired results. Accordingly, the above framework attempts to incorporate flexibility into the recovery system architecture to curb impact of risks and control them.

CONCLUSIONS & FUTURE RESEARCH DIRECTIONS

Emerging sustainable practices and growing green concerns are encouraging the concept of value retrieval through various recovery options such as reuse, remanufacture, recycle, refurbish etc. In this context, organizations are adopting product returns system for managing their returns in order to gain competitive advantage in the market as well as to comply with imposed regulations. Managing product returns at a societal level in an effective and cost-efficient way will help develop sustainable economies. Additionally, CE offers great potential to cost reduction, revenue generation, customer retention and value addition. Accordingly, it is essential for the organizations to ensure the efficient functioning of return channel in order to accomplish their overall goals. In order to so, it is required to curb the risks associated with CE, as they can disrupt the flow of returns and hinder the recovery system performance. These risks and disruptions are a result of various uncertainties and vulnerabilities existing in the return system, which can be due to the changing quantity, quality, time and type of returns. Flexibility can be an effective way to handle the ambiguities, as it enables a system to quickly respond to changing environment. Consequently, in order to handle the uncertainties and complexities of returns, it is necessary to implement flexibility in CE.

In the literature, prior studies have not explicitly investigated the risk aspects of product returns system. To address this gap, the paper has attempted to identify the various risks present in return system under the categories of return collection risk, inspection & disposition risk and reprocessing risk. Further, the study has introduced a flexibility based framework to show the importance of various categories of flexibility in mitigating the recognized risks. Building upon the literature of supply chain risk management and flexibility, the framework presents how each flexibility category is assisting the CE to respond to various identified risks. Accordingly, the research can guide managers in mitigating the CE risks and thus, enhancing its performance and efficiency. Additionally, it will also illustrate the practical implication of deploying flexibility in return channel. The study is of the earliest works closing the links of CE, risk mitigation and flexibility.

The proposed tabular framework can lay the groundwork for further research in the direction of risk management and flexibility deployment in CE. The limitation of the study is that the framework has been not empirically validated. Accordingly, surveys can be conducted to empirically analyze the relationships between the return channel flexibility and risks presented in the framework. Further, various other dimensions of flexibility can be considered to investigate their influence on the risks. Multi attribute decision making techniques can be utilized to quantify the relationship among the risks and various types of flexibility.

REFERENCES

Agrawal, S., Singh, R. K., & Murtaza, Q. (2014). Forecasting product returns for recycling in Indian electronics industry. *Journal of Advances in Management Research,* 11 (1),102-114.

Bai, C., & Sarkis, J. (2013). Flexibility in reverse logistics: a framework and evaluation approach. *Journal of Cleaner Production,* 47, 306-318.

Barker, T. J., & Zabinsky, Z. B. (2011). A multicriteria decision making model for reverse logistics using analytical hierarchy process. *Omega*, *39*(5), 558-573.

Blackburn, J. D., Guide Jr, V. D. R., Souza, G. C., & Van Wassenhove, L. N. (2004). Reverse supply chains for commercial returns. *California management review*, *46*(2), 6-22.

Choudhary, D. & Madaan, J. (2013). Hierarchical decision modeling approach for risk prioritization in sustainable supply chain. *In the proceedings of International Conference on Humanitarian Logistics*, IIM Raipur, India.

Choudhary, D., Madaan, J., & Narain, R. (2014). An Integrated Decision Model for Selection of Third Party Recovery Facilitator (3PRF) for Product Recovery Operations. *Journal of Operations and Supply Chain Management,* 7 (2), 79-101.

Choudhary, D., Shankar, R., & Choudhary, A. (2020). An integrated approach for modeling sustainability risks in freight transportation systems. *Risk Analysis, 40*(4), 858–883.

Choudhary, D., Choudhary, A., Shankar, R., & Hicks, C. (2021). Evaluating the risk exposure of sustainable freight transportation: a two-phase solution approach. *Annals of Operations Research*, 1-35.

Farahani, S., Otieno, W. & Barah, M., (2019). Environmentally friendly disposition decisions for end-of-life electrical and electronic products: The case of computer remanufacture. *Journal of Cleaner Production*, *224*, 25-39.

Fleischmann, M., Bloemhof-Ruwaard, J. M., Dekker, R., van der Laan, E., van Nunen, J., & Van Wassenhove, L. N. (1997). Quantitative models for reverse logistics: a review. *European Journal of Operational Research,* 103, 1–17

Gandhi, S. J., Gorod, A., & Sauser, B. (2012). Prioritization of Outsourcing Risks from a Systemic Perspective. *Strategic Outsourcing: An International Journal,* 5 (1), 39-71.

Gobbi, C. (2011). Designing the reverse supply chain: the impact of the product residual value. *International Journal of Physical Distribution & Logistics Management,* 41 (8), 768-796

Guide, V., & Van Wassenhove, L. (2002). The Reverse Supply Chain. *Harvard Business Review*, 80 (2), 25-26

Jaber, M.Y., & El Saadany, A.M.A. (2009). The production, remanufacture and waste disposal model with lost sales. *International Journal of Production Economics,* 120,115–124

Janse, B., Schuur, P., & Brito, M. P. (2010). A reverse logistics diagnostic tool: the case of the consumer electronics industry. *International Journal of Advance Manufacturing Technology,* 47, 495-513

Koste, L. L., Malhotra, M. K., & Sharma, S. (2004). Measuring dimensions of manufacturing flexibility. *Journal of Operations Management*, 22, 171–196

Krikke, H., Hofenk, D., & Wang, Y. (2013). Revealing an invisible giant: A comprehensive survey into return practices within original (closed-loop) supply chains. *Resources, Conservation and Recycling,* 73, 239–250

Meade, L., & Sarkis, J.(2002). A conceptual model for selecting and evaluating third-party reverse logistics providers. *Supply Chain Management: An International Journal*, 7 (5), 283 –295

Madaan,J., & Wadhwa,S. (2007). Innovative framework for flexible decisions in sustainable enterprise system*. Global Journal of Flexible Systems Management,* 8, 23–34

Madaan J., Kumar, P & Chan Felix T. S. (2012). Decision and Information Interoperability for Improving Performance of Product Recovery Systems. Decision Support Systems, 53 (3), 448–457

Prahinski, C., & Kocabasoglu, C. (2006). Empirical Research Opportunities in Reverse Supply Chains. *Omega,* 34, 519 – 532

Rogers, D. S. & Tibben-Lembke, R.S. (1999). Going Backwards: Reverse Logistics Trends and Practices. Reno, NV: *Reverse Logistics Executive Council*

Rogers, D.S. & Tibben-Lembke, R.S. (2001). An examination of reverse logistics practices. *Journal of Business Logistics,* 22, 129-148

Sawhney, R., (2006). Interplay between uncertainty and flexibility across the value-chain: Towards a transformation model of manufacturing flexibility. *Journal of Operations Management*, 24, 476–493

Singh, R.K. & Agrawal, S., (2018). Analyzing disposition strategies in reverse supply chains: fuzzy TOPSIS approach. *Management of Environmental Quality: An International Journal*, *29*(3), 427-443.

Skinner, L. K., Bryant, P.T., & Richey, R.G. (2008). Examining the impact of reverse logistics disposition strategies. *International Journal of Physical Distribution & Logistics Management,* 38 (7),518-539.

Srivastava, S.K., & Srivastava, R.K. (2006). Managing Product Returns for Reverse Logistics. *International Journal of Physical Distribution & Logistics Management,* 36 (7), 524- 540

Srivastava, S.K. (2008). Network design for reverse logistics, *Omega,* 36, 535–548

Stock, J. R. (2001). The 7 Deadly Sins of Reverse Logistics. *Material Handling Management,* 56 (3), 5-11

Tan, A.W.K. & Kumar, A. (2006). A decision-making model for reverse logistics in the computer industry. *International Journal of Logistics Management,* 17 (3), 331-54.

Tang, C., & Tomlin, B. (2008). The Power of Flexibility for Mitigating Supply Chain Risks. *International Journal Production Economic,* 116,12–27.

Ullah, M. & Sarkar, B., (2020). Recovery-channel selection in a hybrid manufacturing-remanufacturing production model with RFID and product quality. *International Journal of Production Economics*, *219*, 360-374.

Wadhwa, S., Saxena, A. & Chan, F. T. S. (2008). Framework for flexibility in dynamic supply chain management. *International Journal of Production Research*, 46 (6), 1373–1404

Wadhwa, S., Madaan, J., & Chan, F. T. S. (2009). Flexible decision modeling of reverse logistics system: A Value Adding MCDM Approach for alternative selection. *International Journal of Robotics and Computer-Integrated Manufacturing,* 25 (2), 460–469

DECISION SCIENCES INSTITUTE
Data-driven supply chain and firm performance: Role of risk management capabilities in global pandemics

Attique ur Rehman
PhD Candidate
Suleman Dawood School of Business
Lahore University of Management Sciences, Pakistan
Email: attique.rehman@lums.edu.pk

Muhammad Shakeel Sadiq Jajja
Associate Professor
Suleman Dawood School of Business
Lahore University of Management Sciences, Pakistan
Email: ssj@lums.edu.pk

ABSTRACT

Purpose: Recent technological advancements have made firms able to analyze data for performance improvements. Drawing on technology-organization-environment (TOE) context, this study explains the effect of the data-driven supply chain on supply chain performance through supply chain risk management capabilities of robustness and resilience in the wake of global pandemics such as Covid-19

Design/Methodology/Approach: This paper uses data from 200 manufacturing firms from Pakistan. Structural equation modeling is used for mediation analysis of supply chain risk management capabilities towards data-driven supply chain and firm supply chain performance.

Originality/Value: The paper extends debate in supply chain literature by providing an overarching view of the data-driven supply chain and how it affects performance especially in the context of a global pandemic. Firms use data analytics to improve their supply chain resilience and robustness that in turn affect supply chain performance. This mechanism helps the firm to understand the way they may use and analyze enormous data being generated in their supply chains to achieve superior performance by developing their risk management capabilities.

Research and practical Implications: Firms can use data-driven supply adoption to manage supply chain risks effectively.

KEYWORDS: Data-driven supply chain, supply chain resilience, supply chain robustness, supply chain performance, developing countries

INTRODUCTION

Supply chain risk management has become an area of sheer importance for academics and practitioners, especially when the risks are disruptive in nature (Tang, 2006; Craighead et al., 2007; Govindan and Fattahi, 2017; Ivanov, 2020). While operational risks are concerned with day-to-day turbulences in the supply chain such as a change in lead-time, and supply and demand fluctuations, the disruption risks are driven by natural disasters. These disasters include tsunami, earthquakes, and man-made catastrophes such as terrorist attacks. Spurred by highly publicized and well-documented events such as terrorist attacks e.g., New York, 2001, London 2005 and Jakarta 2009, Tsunami, 2004, 2011, diseases e.g., SARS, 2003, avian/bird flu, 2005 and swine flu, 2009, the interest in supply chain risk issues has steadily grown. Epidemic outbreaks that escalate to the level of pandemics are a special case of supply chain risk that are characterized by long-term disruption, ripple effect, and high level of supply chain uncertainty (Ivanov, 2020).

The issues that expose firms to supply chain risks can be managed by visibility and information sharing across supply chains for decision making (Urciuoli and Hintsa, 2018). With the inception of industry 4.0, the internet of things, big data, and other digital technologies, firms have shifted their focus towards data analytics to drive capabilities, manage risks and achieve performance improvements (Govindan et al., 2018). Academics have referred to this phenomenon as data-driven supply chain (DDSC). Data analytics or DDSC enhances firms' supply chain capabilities (Yu et al., 2018) that are articulated in several dimensions such as supply chain integration (Flynn et al., 2010), supply-side flexibility (Wagner et al., 2018), information exchange and process coordination and responsiveness (Yu et al., 2018). In addition, data analytics in supply chains have contributed in terms of logistics management and planning (Tan et al., 2015; Zhong et al., 2015), optimization of inventory policy (ÇAkıCı et al., 2011), waste minimization (Mishra and Singh, 2018), production planning and scheduling (Zhong et al., 2015) and financial performance (Yu et al., 2018). However, in this information era, data-driven supply chain may also be useful towards enhancing firms' supply chain risk management capabilities, especially when risk is triggered through disruptions in the environment. Further, the use of data becomes prevalent in the context of uncertain circumstances that arise in a global pandemic scenario.

Data-driven supply chain (DDSC) refers to the information sharing across the entire supply chain to connect supply chain partners and provide end-to-end supply chain data access (Sanders, 2014). Wamba and Akter (2019) argued that firms need management, talent and technological capabilities to implement a data-driven supply chain. Management and talent capabilities refer to the top management commitment and personnel capabilities (Gunasekaran et al., 2017), corporate strategy (McAfee et al., 2012), a data-driven culture (Dubey et al., 2019), and readiness to invest in new technologies (Akter et al., 2016). Technology capabilities refer to readiness in technology adoption (Ransbotham et al., 2015), data resources (Barton and Court, 2012), technology capability (Akter et al., 2016), and information sharing (Gunasekaran et al., 2017). These capabilities are aligned with the technology-organization-environment (TOE) context. TOE framework was proposed by Tornatzky et al. (1990) that enables the adoption decisions regarding technological innovation by firms based on the three aforementioned factors. TOE has been used in the literature towards innovation adoption, e-business, and e-procurement decisions, cloud computing, and RFID technologies (Oliveira and Martins, 2011; Oliveira et al., 2014; Akter et al., 2016) but using TOE context as an enabler of data-driven supply chain and supply chain risk management capabilities has yet to be received attention in supply chain literature.

Supply chain risk management is viewed as a holistic phenomenon that incorporates a set of capabilities, a firm develops to enhance its performance (Riley et al., 2016). These capabilities include robustness, resilience, and agility (Wieland and Marcus Wallenburg, 2012; Kwak et al., 2018) that help firms manage risk effectively (Colicchia and Strozzi, 2012). Based on the replication and imitation of the dynamic capabilities of Teece et al. (1997), we argue that a data-driven supply chain, as a dynamic capability, enhances firms' SC risk management capabilities in terms of robustness and resilience. The concept of robustness in supply chain risk setting is the ability of supply chains to remain effective for all plausible future scenarios (Klibi et al., 2010) whereas supply chain resilience is the ability to move towards its original state or move to a new desirable state after being disturbed (Christopher and Peck, 2004). We aim to answer following research questions: "How the TOE context enables firm's data driven supply chain adoption?", and, "What is impact of data driven supply chain on supply chain performance through supply chain risk management capabilities?" This paper theorizes Technology-Organization-Environment (TOE) context to establish the motivation of firms to adopt a data-driven supply chain. In addition, we argue that supply chain risk management mediates the relationship between DDSC and supply chain performance.

Drawing on the TOE and DCV approach, we make some significant contributions to the literature. First, we inform supply chain literature by making use of TOE as an enabler towards DDSC adoption by firms. Second, based on DCV, we argue that DDSC adoption allows the firm to develop risk management capabilities in its supply chain in terms of SC resilience and robustness. Third, we explain the explain the mediation effect of supply chain risk management towards DDSC and supply chain performance. This provides an overarching mechanism that allows firms to develop a competitive advantage in their supply chain through technological advancements and effective risk management.

The rest of the paper is organized as follows. Section 2 comprises of literature review and hypotheses development. The research methodology is presented in section 3 including instrument design, sample, data collection, and methods used. Section 4 reports the key findings followed by discussion, managerial implications, and future research directions in sections 5 and 6 respectively.

Background

Technological, organizational, and environmental (TOE) framework

The TOE framework developed by (Tornatzky et al., 1990) presents dimensions of an enterprise's context that affect the adoption and implementation of innovations. It posits that various contextual factors affecting an organization's ability to adapt and implement technological innovation stem from technological, organizational, and environmental facets of a firm's context (Chen et al., 2015). TOE framework assumes that the adoption process within a firm is effectively established through the right match between external and internal factors (Arpaci et al., 2012). Though TOE does not offer a set of concrete factors that affect the implementation of technology in a firm's operations, rather it provides taxonomies of the factors that affect technology adoption by firms (Ven and Verelst, 2011; Ismail and Ali, 2013). The technological context dimension focuses on how technological practices and structures can influence the adoption process. The organizational context manifests common organizational attributes that support or constrain innovation adoption. These attributes may contain the firm size, scope, managerial characteristics such as decision-making structure, quality of communication, and support by management. Environmental context reflects the stakeholders that surround an enterprise such as suppliers, customers, competitors, the government and community, etc. who determine the need for innovation by the firm (Aboelmaged, 2014).

The TOE framework has been examined across different disciplines and contexts to examine its theoretical strength, empirical support, and usefulness in investigating the readiness, adoption, and deployment of various forms of innovation, such as the adoption of complex information system innovations, information and communication technology, cloud computing, e-commerce, and ERP systems in SMEs in England (Ghobakhloo et al., 2011; Kuan & Chau, 2001; Lian et al., 2014; Srivastava & Teo, 2010). Similarly, TOE has been used in supply chain management literarure to study the adoption of digital manufacturing technologies, web 2.0, and big data analytics adoption (Chen et al., 2015; Tarofder et al., 2019; Gillani et al., 2020). Table 1 summarizes a representative set of studies using the TOE framework in general and particularly, in supply chain literature.

Author and Year	Journal	Paper type/Methods	Empirical context (Country/Industry)	Theory	Key pertinent findings
Orji et al. (2020)	IJPE	Empirical quantitative	UK, India/Freight logistics industry	Human-Organization-Technology (HOT)	• Competitive pressure, customer satisfaction, and affordability are significant factors towards the use of social media in the social sustainability of supply chains
Wong et al. (2020)	International journal of information management	Empirical quantitative	Malaysia/SMEs	--	• Competitive pressure, complexity, cost, and relative advantage are significant towards the behavioral intention of blockchain adoption
Tarofder et al. (2019)	Industrial engineering and management systems	Empirical quantitative	Malaysia/Manufacturing	--	• TOE significantly affects web 2.0 adoption • Top management support is of utmost important in web 2.0 technology adoption in the firm's supply chain
Mahroof (2019)	International journal of information management	Empirical qualitative	United Kingdom/manufacturing	Unified theory of acceptance and use of technology	• TOE affects artificial intelligence adoption in warehouse and logistics supply chain • TOE provides opportunities to firms to develop artificial intelligence systems to create a competitive advantage
Lai et al. (2018)	IJLM	Empirical quantitative	China/Logistics	Innovation diffusion theory	• Perceived benefits, top management support, competition, government policies, and supply chain connectivity influence adoption intention of big-data analytics
Pool et al. (2015)	International journal of logistic systems and management	Empirical quantitative	Iran/Manufacturing	--	• TOE is significant towards RFID acceptance and adoption in manufacturing firms

Sugathan et al. (2015)	Advances in intelligent systems and computing	Empirical quantitative	Malaysia/manufacturing	Input-output theory	• TOE is a driver of green supply chain management practices that leads to positive technological and environmental outcome
Chen et al. (2015)	Journal of management information systems	Empirical quantitative	USA/manufacturing	Dynamic capability theory, information process theory	• Technology readiness, top management support, competitive pressure are significant driver towards big-data analytics adoption
Lian et al. (2014)	International journal of information management	Empirical quantitative	Taiwan/Hospital industry	--	• CIO innovativeness, data security, compatibility, top management support, industry pressure, and adequate resources play a pivotal role in the adoption of cloud computing
Chan and Chong (2013)	IJPR	Empirical quantitative	Malaysia/Manufacturing	Innovation diffusion theory	• TOE enhances mobile SCM diffusion • Technological and organizational factors significantly affect mobile SCM adoption of firms
Shaik et al. (2013)	International journal of information systems and supply chain management	Empirical quantitative	Conceptual	Relational view	• TOE affects information system adoption in supply chains
Sila (2013)	Electronic commerce research	Empirical quantitative	America/Manufacturing	Innovation diffusion theory	• TOE significantly affects B2B electronic commerce adoption
Ifinedo (2011)	Internet research	Empirical quantitative	Canada/SMEs	--	• Perceived benefits, management support, IT competence, IS vendor support and financial support positively influence internet and e-business technologies
Ghobakhloo	Industrial	Empirical	Iran/Manufacturing	--	• E-commerce adoption is affected

et al. (2011)	management and data systems	quantitative	SMEs		by perceived relative advantage, CEO's innovativeness, information intensity, and competition
Oliveira and Martins (2010)	Industrial management and data systems	Empirical quantitative	European Union/telecommunication and tourism	--	• Perceived benefits, technology readiness, competitive pressure, and trading partner collaboration are drivers for e-business adoption
Zhu et al. (2010)	International journal of information management	Empirical quantitative	China/retail industry	--	• Implementation quality, organizational readiness, leadership involvement, competition are critical towards post-implementation ERP success
Srivastava and Teo (2010)	Communications of ACM	Empirical quantitative	Secondary data from 113 countries	--	• ICT infrastructure is significant towards e-business adoption and e-government • The national environment is pivotal towards e-business adoption • E-business and e-government are critical towards enhancing national economic performance
Dwivedi et al. (2009)	Journal of enterprise information management	Empirical quantitative	England/SMEs	--	• Top management support, size, organizational readiness are significant factors towards organization-wide ERP systems
Kuan and Chau (2001)	Information and Management	Empirical quantitative	Hong Kong/Small industries	--	• TOE affects EDI adoption decisions

Table 1: Themes of the TOE framework in the literature

Data-driven supply chain adoption

In supply chain literature, data analytics and big-data have been considered as a means towards a data-driven supply chain (Yu et al., 2018). Big data refers to the data that is in such volume, velocity, variety, and veracity that typical computing infrastructure cannot process it (McAfee and Brynjolfsson, 2008; Lycett, 2013). Data-driven supply chain allows firms to share information across the entire supply chain to connect supply chain partners and provide end-to-end supply chain data access (Sanders, 2014). To be most effective, it should be embedded into the organizational process (LaValle et al., 2011). Data-driven supply chain makes use of RFID data, clickstream data, transaction, video, and voice data over the supply chain for effective decision making (Wamba et al., 2017; Wamba and Akter, 2019). For example, Roßmann et al. (2018) explored the role of data analytics to enhance demand forecast, reduce safety stocks, and improve the overall performance of the supply chain. Extant literature has argued the role of the use of data independently across different functions in a firm's supply chain and by different supply chain partners (Sanders, 2014; Kamble and Gunasekaran, 2020). The use of such data across different functions and by different supply chain partners has limitations as such data has been used in silos for decision making (Yu et al., 2018). For organizations, to maximize their use of information and increase the utility, data must be shared across processes not within the organizations, but also outside the organization, providing end-to-end support across the supply chain (Sanders, 2014). Data-driven supply chain provide the firms with such mechanisms to use the data across end-to-end supply chain and helps firms improve supply chain functionality based on the use of datasets available in different functions of the business such as operations, procurement, and marketing and sales (LaValle et al., 2011). For example, firms improve their demand forecasting, supply planning, and manufacturing operations efficiency using their own data and supplementary data from customers and suppliers such as raw material, inventory, and delivery data.

The existing literature on a data-driven supply chain focuses on the use of data analytics across the supply chain (Gawankar et al., 2020). A structured and data-driven decision making in the supply chain yields better results as compared to unstructured and subjective decision making (Yazdani et al., 2017). The literature reports that firms use only one-fourth of the existing data while making important decisions and less than 1% of information is analyzed, though 80% of the time of data scientists are used productively to enhance the decision-making process (Grover and Kar, 2017).

SCRM capabilities

The supply chain risk management perspective has gained importance for both academics and practitioners (Colicchia and Strozzi, 2012; Golgeci and Ponomarov, 2013). The world has seen a significant disruption in supply chains owing to the developed Covid-19 scenario across the globe (Ivanov, 2020). The main concern of supply risk management is to reactively and proactively manage the firm's supply chains in times of crisis and disastrous situations that are become increasingly prevalent (Manuj and Mentzer, 2008; Rao and Goldsby, 2009; Golgeci and Ponomarov, 2013). Supply chain resilience and robustness can be seen as capabilities that firms develop to mitigate and proactively prepare for risks (Hendricks and Singhal, 2005; Hendricks et al., 2005). Though supply chain literature has conflated resilience and robustness and used them interchangeably (Christopher and Peck, 2004) and/or switched the causal logic, their conceptual meaning is distinct. (Brandon-Jones et al., 2014). Wieland and Marcus Wallenburg (2012) argued that every supply chain must be robust and resilient to manage risk effectively.

The concept of resilience in the supply chain is evolving. Though many definitions of resilience exist across various academic disciplines (Haimes, 2009; Kamalahmadi and Parast, 2016), the core concept of resilience can be viewed as an ability to withstand the effect of disruption. Supply chain resilience is defined as "*the ability of a system to return to its original state, within an acceptable period, after being disturbed.*" (Brandon-Jones et al., 2014). Resistance implies that when a system is affected by the disruption, it has an inbuilt ability to restore to its original state or move to a new improved state. In literature, adjusting to unexpected changes is a hallmark of supply chain resiliency (Krause et al., 2009; Ralston and Blackhurst, 2020). Resilience is an important supply chain risk management strategy (Wieland and Marcus Wallenburg, 2012; Mackay et al., 2019) as a means to address the turbulence generated from uncertain events in the supply chain.

Robustness is frequently discussed in engineering literature (Bai et al., 2016; Zhou et al., 2017), and risk literature (Haimes et al., 1998) and has also emerged in supply chain literature due to increased supply chain vulnerability and uncertainties. Supply chain robustness is defined as *"the ability of the supply chain to maintain its function despite internal or external disruptions."* (Kitano, 2004). This definition implies that the firm must be able to continue its operations despite turbulence and disruptions in the supply chain, resisting the disruption at the same time (Stonebraker et al., 2009). Wieland and Marcus Wallenburg (2012) argued that robustness is a proactive strategy that enables the supply chain to resist change and disruption without adapting its initial stable configuration. Though supply risk literature struggles with definitional clarity of robustness and resilience concepts in supply chains (Aven and Zio, 2014), Saenz et al. (2015) advocated that supply chain robustness is a component of resilience.

Supply chain performance

Supply chain performance is defined as "*the efficient and effective execution of supply chain tasks*." (Autry et al., 2014) and as *"the extent to which the supply chain can meet customer requirements and deliver product on-time"* (Tarafdar and Qrunfleh, 2017). This definition implies the ability of firms to manage all three dimensions of the supply chain i.e., supply, operations/production, and demand. Many firms look at continuous improvement in their supply chains as a tool to evaluate performance. These improvements are measured in terms of speed, flexibility, customer satisfaction, delivery performance, and cost (Beames, 2004). As compared to other performance measures such as financial performance, operational performance, and customer performance, there is a growing interest of academics and practitioners in the supply chain performance of firms (Maestrini et al., 2017). Supply chain performance plays a critical role in organizations because attaining such performance has been rendered more difficult by the increased complexility of the context in which firms operate (Wamba et al., 2020).

TOE and Data-driven supply chain adoption

TOE framework has been proved to be the most widely used and prominent framework towards technology adoption (Hsu et al., 2014; Lai et al., 2018). In literature TOE framework has been used for Electronic Data Interchange (EDI) adoption (Kuan and Chau, 2001), ERP system implementation and success (Dwivedi et al., 2009; Zhu et al., 2010), e-commerce adoption (Ghobakhloo et al., 2011) and big data analytics adoption (Lai et al., 2018). The rich application of TOE towards multilevel determinants of technology adoption validate its use towards data-driven supply chain adoption of firms.

Technological context focuses on the factors that become enablers towards the decision making of new technology adoption (Tornatzky et al., 1990). The technology context is a combination of technology infrastructure and technology competence (Aboelmaged, 2014). Technology infrastructure refers to the shared platform (e.g., computer hardware, software, networking technologies) that are pivotal to implement new technology solutions throughout the organization (Bhattacherjee and Hikmet, 2008). Beyond physical technological assets, the intangible technical knowledge of human resources offers the technical know-how needed to develop and diffuse the technology in a firm's supply chain operations (Zhu and Kraemer, 2005). Technological competence refers to the expertise a firm has in information and communication technologies and related e-business activities (Aboelmaged, 2014).

The availability of well-designed technological infrastructure helps organizations to infuse better decision making in their supply chains (Chavez et al., 2017). In contrast, lack of such infrastructure may limit the organization's ability towards improved decision making and have a holistic view of information exchange, coordination, and integration in the supply chain (Armstrong and Sambamurthy, 1999; Bhattacherjee and Hikmet, 2008; Chavez et al., 2017). Research has shown that firms with technology infrastructure have increased their readiness for data-driven technology implementation in supply chains (Aboelmaged, 2014; Chavez et al., 2017). Similarly, a growing body of knowledge has identified the need and importance of technology competence towards the adoption of innovative practices. Ifinedo (2011) argued that when firms are technologically competent, they are more inclined to effectively implement technological practices in their supply chain operations (Chavez et al., 2017; Yu et al., 2018). Ranganathan et al. (2011) considered managerial IT knowledge as one of the key antecedents towards web-enabled supply chain management. Likewise, Lin and Lee (2005) claimed that the firms whose employees have necessary skills and technical knowledge are more likely to develop and implement e-business applications. Thus we hypothesize:

H1 Technology context of firms positively affects data-driven supply chain adoption

Organizational factors play a pivotal role in data-driven technology adoption in firms (Chen et al., 2015; Lai et al., 2018). Former studies identify top management support, organizational readiness, and employees' expertise as some key characteristics of organizations' adoption of information technology (Lai et al., 2018). Prior literature suggests that when top managers are convinced with the benefits of the data across supply chain and have positive beliefs towards the use of such data-driven systems for organizational competitiveness, they will take actions to deploy such systems (Liang et al., 2007). Top management support can only be effective if the firm has available resources and infrastructure (Cooper and Zmud, 1990). Particularly, prior research suggests that these resources include financial capital and technology sophistication (Li, 2009). Technology sophistication captures not only infrastructure but also human resources that are used to deploy the technology infrastructure and manage data-driven supply chains. Chen et al. (2015) suggest that the availability of professionals with the skills and capacities to handle data develops organizational readiness and helps firms deploy data-driven supply chains effectively. Hsu et al. (2014) suggest that firms with stronger technological capabilities will demonstrate a higher capability to adopt cloud services. Similarly, firm size and available resources positively affect the adoption of RFID technologies by firms (Wamba et al., 2017). In this paper, we focus on top management support and availability of key personnel as specific components of organizational context that affect data-driven supply chain adoption in firms. Due to the support from internal organizational factors, we propose:

H2 Organizational context of firms positively affects data-driven supply chain adoption

In the TOE framework, environmental context is defined as "the climate in which an organization conducts its operations." (Maduku et al., 2016). This 'climate' includes factors such as competitive pressure, government legislations, supply chain partner pressure, technology standards, and consumer readiness. Pressure from competitors refers to the pressure firms feel from their competition while operating in a certain industry (To et al., 2006). Competitive pressure pushes the firm to adopt new technologies and make changes in their supply chain mechanisms (Obal, 2017). Early technology adopters enjoy a first-mover advantage in a particular industry. Firms tend to adopt the latest technological advancements to boost their competitive advantage when faced with severe competition (Zhu et al., 2006; Sun et al., 2020). Similarly, firms recognize the importance of collaboration with external supply chain partners to improve their performance (Cao and Zhang, 2011). To successfully leverage the resources and knowledge across the supply chain, firms must have systems that are compatible with those of their trading partners (Sun et al., 2020) such as RFID tags for inventory control (Lai et al., 2006). For building a supply chain partnership effectively, firms must keep up to the pace of the partner's use of new technologies. Based on the arguments, we propose the following hypothesis:

H3 Environmental context of firms positively affects data-driven supply chain adoption

Dynamic capability view

Dynamic capability view (DCV) helps the firm identify the sources of competitive advantage in changing business environments (Teece and Pisano, 1998; Zollo and Winter, 2002). DCV argues the realization of sustainable competitive advantage that depends on firms' ability to develop key competencies in their supply chains such as risk management (Kilubi and Haasis, 2015; Kilubi and Rogers, 2018). It has been argued that knowledge transfer mechanisms can help build dynamic capabilities and core competence in a firm's supply chain (Tsou et al., 2014) in an uncertain environment. Drawing on DCV, we define a data-driven supply chain as a holistic process that provides robust insights for real-time decision making through complimentary use of technological and personnel resources. In the context of DCV, the ability to leverage data can be considered as a firm's capability (Marchand et al., 2000) as it helps the firm develop supply chain strategies such as supply chain such as coordination, integration, and responsiveness (Yu et al., 2018). It is argued that firms need superior resources to enhance data-driven capabilities. Yu et al. (2018) used the concept of complementary assets by Teece et al. (1986) that stems from a resource-based view of Wernerfelt (1984) to explain the role of data-driven supply adoption towards firm's financial performance. Resources involved in forming complementary assets may be physical, human, or organizational (Barney, 1991). 'Technology' and 'organization' from the TOE framework provide ample basis for complementary assets that help firms to adopt a data-driven supply chain. Thus, DDSC adoption is a dynamic capability that firms develop and it can help enhance supply chain risk capabilities.

Based on a resource-based perspective, supply chain resilience and robustness can be seen as dynamic capabilities that firms develop to mitigate risks (Hendricks and Singhal, 2005; Hendricks et al., 2005). These capabilities are reflected in terms of supply chain resilience and robustness. DCV lens argues that supply chain resilience is the ability of a firm to bounce back to its original state after encountering a disruption (Abeysekara et al., 2019). Simiarly, supply chain robustness deals with the shock absorption capability of a firm after being exposed to disruptive risk in supply chain. Both, supply chain resilience and robustness capabilities are pronounced based on the use of data-driven supply chain adoption. The next part builds this argument.

DDSC adoption and SCRM capabilities

Teece et al. (1997) argued that dynamic capabilities are replicated and imitated and transferred in a firm's processes. The concept of replication involves transferring or redeploying core competencies from one setting to another. Based on this argument, firms can use data analytics and DDSC as a dynamic capability to inform different supply chain functions such as purchasing, production and operations, distribution, and after-sales services (Hopkins and Brokaw, 2011; Sanders, 2014; Yu et al., 2018) to enhance competitive advantage in these functions. Using real-time data, firms can manage their demand, supplies, reducing defects, and adjust to uncertain changes in the supply chain. Bringing technology and integrating information into the supply chain help firms to heighten risk management capabilities (Kwak et al., 2018). For example, the application of integrated communication systems across the supply chain facilitates enhanced robustness and resilient capabilities across the supply chain (Waters, 2011). Kwak et al. (2018) argued that supply chain innovation through enhanced information channels and improved breadth and speed of information sharing across the supply chain positively affects supply chain risk management capabilities of robustness and resilience. Similarly, Riley et al. (2016) explored the role of internal supply chain integration, information sharing and human resource training towards the warning and recovery capability of firms based on Information sharing across the supply chain. To continue improving supply chain effectiveness, firms need to leverage large supply chain datasets and a mechanism to use these datasets through superior technological advancements across the supply chain. Overall DDSC enables firms to achieve dramatic improvements in managing complex and risk-exposed supply chains.

Thus DDSC can develop superior capabilities to handle and manage supply chain risks through resilience and robustness, so it is hypothesized that:

H4 DDSC adoption positively affects SC robustness capability.

H5 DDSC adoption positively affects SC resilience capability.

“A resilient supply chain is certainly robust” and “a resilient supply chain must also be adaptable”. (Christopher and Rutherford, 2004). Similarly, Wieland and Marcus Wallenburg (2012) argued that agility and robustness are dimensions of resilience. On one hand, robustness is considered as a built-in feature of a resilient supply chain, whereas on the other hand, robustness and resilience have also been treated as independent supply chain risk management capabilities (Riley et al., 2016; Kwak et al., 2018). SC robustness is seen as a shock absorber to withstand disruptions at a tolerable level (Kwak et al., 2018). Further, SC robustness is argued as a reactive risk mitigation strategy that helps firm manage supply chain risk effectively. Proactive risk management in supply chain leads reactive risk mitigation (Gouda and Saranga, 2018)Whereas, SC resilience is seen as a proactive strategy that helps firm foresee risk through flexibility, anticipation, and risk preparation (Klibi and Martel, 2012).The role of SC robustness is pivotal during the initial stages of disruption where risk mitigtion can minimize or even eliminate the regular occurrence of risk (Tang, 2006), whereas supply chain resilience is critical because of its proactive nature to anticipate the impact of unexpected event or disruption in the supply chain. Adaptability is the key component of SC resilience along with responsiveness to return to the original or new state of the supply chain. (Ponomarov and Holcomb, 2009). We argue that SC resilience will also enhance SC robustness towards effective SC risk management. Thus, we propose:

H6: SC resilience positively affects SC robustness.

DDSC, SCRM capabilities, and SC performance

Prior studies have identified the role of use of data-driven and technological capabilities of the firms towards superior performance. For example, Wamba et al. (2017) argued that big-data analytics capability positively affects firm’s financial and market performance. Similarly, Chen et al. (2015) found that the data-driven capability of firms has a significant effect on performance because real-time data leverage the ability of firm to make better decisions thus enhance performance. Also, blockchain technology adoption has a significant positive impact on supply chain performance (Wamba et al., 2020) as it makes data transparent across the supply chain to be accessed by different supply chain partners. This adds brings synergy in the decision making and enhances performance. Literature has stressed the importance of knowledge exchange and logistics innovation towards supply chain performance (Wamba et al., 2020). The firms capable of big-data orientation have the potential to revolutionize supply chain performance (Waller and Fawcett, 2013; Chavez et al., 2017) due to their capability of dealing with data in real-time and making better decisions for the supply chain. Successful exploitation of big data is essential for nurturing business performance (Kwon et al., 2014). The use of a data-driven supply chain allows the firm to harness business opportunities and achieve superior performance. Thus, we argue that:

H7 DDSC positively affects SC performance of firms.

Supply chain risk management positively affects firm performance (Wagner and Bode, 2008; Manuj et al., 2014; Kauppi et al., 2016). As discussed, supply chain risk management capabilities have been bifurcated as SC resilience and SC robustness capabilities. A resilient supply chain alleviates the firm's capability to not only absorb the risk but also to speedily recover and return to its original state. This can affect performance. (Pettit et al., 2013). The longer it takes for the supply chain to return to its original state, the greater would be the damage (Gunasekaran et al., 2015). Prior literature has argued that SC resilience is positively associated with firm performance i.e., market growth, market share and profitability (Abeysekara et al., 2019). It is argued that supply chain performance is also impacted by supply chain resilience, thus we hypothesize:

H8 SC resilience affects SC performance of firms.

Supply chain robustness is an important indicator of supply chain performance (Dong and Informatics, 2006). Raj and Srinivas (2019) argued that SC robustness has a positive impact on performance as it enhances the ability of firms to withstand the exposed risk and serve as a shield to minimize the impact of risks on supply chain activties. Extant literature studied the role of supply chain robustness towards performance (Saenz et al., 2015; Kwak et al., 2018; Madzimure and Management, 2020). The negative impacts of supply chain disruptions on supply chain performance (Hendricks and Singhal, 2005) can be eliminated by supply chain robustness. Thus, hypothesize:

H9 SC robustness affects SC performance of firms.

Mediation effects

It is argued that data-driven capabilities in supply chains can allow firms to manage risk effectively that further enhance performance (Kshetri, 2018; Wamba et al., 2020). SC resilience and DCV literature mentioned that firms operating in highly dynamic environments need to coordinate, integrate, and configure their resources according to changing business environment to achieve competitive advantage (Teece, 2014). Building data-driven supply chain orientation is one of the ways to develop synergistic effects across the supply chain and enhance performance through effective supply chain risk management. Further, SC resilience and robustness have been argued as mediators towards supply chain innovation and financial performance in manufacturing firms in Thailand (Srimarut and Mekhum, 2020). Based on the phenomenon of replication and imitation of dynamic capabilities Teece et al. (1997), it is argued that SC risk management capabilities in terms of SC resilience and robustness enhance the impact of DDSC on firm supply chain performance. DDSC allow firms to develop risk management capabilities through real-time data analyses across supply chain that is reflected in enhanced supply chain performance. Following these lines of arguments, and building on H4 and H5, we hypothesize

H10 SC robustness mediates the relationship between DDSC and SC performance.

H11 SC resilience mediates the relationship between DDSC and SC performance.

Information sharing and the use of data across the supply chain helps firms develop appropriate mitigation strategies and further develop proactive strategies. It is argued that reactive capabilities are developed before proactive capabilities. (Morash, 2001). A data-driven supply chain helps the firm develop resilience in their supply chain that further leads to robustness and affect performance. It is argued that:

H12: SC resilience mediates the relationship between DDSC and SC robustness.

Supply chain resilience has long-term effects on the supply chain as compared to robustness (Kwak et al., 2018) as it is developed as a proactive capability to manage supply chain risk. From this perspective, a distinction is drawn between SC robustness and resilience where the former builds capacity in terms of adaptation, reaction, and growth when faced with severe disruption, the latter allows the supply chain to react to and absorbs the risk during initial stages of disruptions. (Riley et al., 2016; Kwak et al., 2018). Before developing a reactive capability, supply chain needs

a proactive capability risk management capability to enhance SC performance. Thus, it is proposed that:
H13: SC robustness mediates the relationship between SC resilience and SC performance.

Methodology
In this research, to validate our conceptual model, the survey data from manufacturing sector of Pakistan is used. We focus on four main industrial sectors (Table 6). Pakistan economy is dominated by manufacturing sector as compared to services sector as it contributes to 14% to overall GDP of Pakistan. (Bank, 2019). Further, the exposure of manufacturing sector towards supply chain risk in this global pandemic has also sensitized the operational capability of different production and logistics units in Pakistan.

A total of 300 companies from different sectors were selected to participate in the survey and a total of 250 surveys were collected. After filtering, a total of 200 usable responses were screened as shown in Table 2. The original questionnaire was drafted in English and was further translated by the researcher in case the respondent had issues in understanding the language. Active involvement of the managers ensured the relevance of the instrument and its content validity. The respondents were the employees who relate to the operations decisions of the organization (operations managers, supply chain managers, plant managers, etc.). Potential respondents were contacted through the research team and questionnaires were sent to the organizations through email, or ordinary mail.

Industries

Firm Size (Number of employees)	**Expected data in percentage**
Small (<250)	45 – 50%
Medium (250 – 500)	30 – 40%
Large (> 500)	10 – 25%
Apparel and Textile • Weaving and warping • Dyeing and printing • Finishing • Stitching	**10 – 15%**
FMCG • Food manufacturers • Beverages • Hygiene product manufacturers	**20 – 25%**
Healthcare • Pharmaceuticals • Surgical • Paper and plastic healthcare products (facemasks, gloves, etc.)	**15 – 20%**
Automobiles • Assemblers • Spare part manufacturers	**10 – 15%**

BIBLIOGRAPHY

Abeysekara, N., Wang, H., & Kuruppuarachchi, D. J. B. P. M. J. (2019). Effect of supply-chain resilience on firm performance and competitive advantage.

Aboelmaged, M. G. (2014). Predicting e-readiness at firm-level: An analysis of technological, organizational and environmental (TOE) effects on e-maintenance readiness in manufacturing firms. *International Journal of Information Management, 34*(5), 639-651.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics, 182*, 113-131.

Armstrong, C. P., & Sambamurthy, V. (1999). Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Information systems research, 10*(4), 304-327.

Arpaci, I., Yardimci, Y. C., Ozkan, S., & Turetken, O. (2012). Organizational adoption of information technologies: A literature review. *International Journal of ebusiness and egovernment Studies, 4*(2), 37-50.

Autry, C. W., Williams, B. D., & Golicic, S. J. J. o. B. L. (2014). Relational and process multiplexity in vertical supply chain triads: An exploration in the US restaurant industry. *35*(1), 52-70.

Aven, T., & Zio, E. J. R. a. (2014). Foundational issues in risk assessment and risk management. *34*(7), 1164-1172.

Bai, Y.-N., Huang, N., Wang, L., & Wu, Z.-X. J. S. r. (2016). Robustness and vulnerability of networks with dynamical dependency groups. *6*, 37749.

Bank, W. (2019). World development indicators.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management, 17*(1), 99-120.

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard business review, 90*(10), 78-83.

Beames, S. (2004). Critical elements of an expedition experience. *Journal of Adventure Education & Outdoor Learning, 4*(2), 145-157.

Bhattacherjee, A., & Hikmet, N. (2008). Reconceptualizing organizational support and its effect on information technology usage: Evidence from the health care sector. *Journal of Computer Information Systems, 48*(4), 69-76.

Brandon-Jones, E., Squire, B., Autry, C. W., & Petersen, K. J. J. J. o. S. C. M. (2014). A contingent resource-based perspective of supply chain resilience and robustness. *50*(3), 55-73.

ÇAkıCı, Ö. E., Groenevelt, H., & Seidmann, A. (2011). Using RFID for the management of pharmaceutical inventory—system optimization and shrinkage control. *Decision Support Systems, 51*(4), 842-852.

Cao, M., & Zhang, Q. J. J. o. o. m. (2011). Supply chain collaboration: Impact on collaborative advantage and firm performance. *29*(3), 163-180.

Chan, F. T., & Chong, A. Y.-L. J. I. J. o. P. R. (2013). Determinants of mobile supply chain management system diffusion: a structural equation analysis of manufacturing firms. *51*(4), 1196-1213.

Chavez, R., Yu, W., Jacobs, M. A., & Feng, M. (2017). Data-driven supply chains, manufacturing capability and customer satisfaction. *Production Planning & Control, 28*(11-12), 906-918.

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems, 32*(4), 4-39.

Christopher, M., & Peck, H. (2004). Building the resilient supply chain. *The international journal of logistics management, 15*(2), 1-14.

Christopher, M., & Rutherford, C. J. C. e. (2004). Creating supply chain resilience through agile six sigma. *7*(1), 24-28.

Colicchia, C., & Strozzi, F. (2012). Supply chain risk management: a new methodology for a systematic literature review. *Supply Chain Management: An International Journal, 17*(4), 403-418.

Cooper, R. B., & Zmud, R. W. J. M. s. (1990). Information technology implementation research: a technological diffusion approach. *36*(2), 123-139.

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., & Handfield, R. B. (2007). The severity of supply chain disruptions: design characteristics and mitigation capabilities. *Decision Sciences, 38*(1), 131-156.

Dong, M. J. I. J. o. S. O., & Informatics. (2006). Development of supply chain network robustness index. *1*(1-2), 54-66.

Dubey, R., Gunasekaran, A., Childe, S. J., Blome, C., & Papadopoulos, T. (2019). Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. *British journal of management, 30*(2), 341-361.

Dwivedi, Y. K., Papazafeiropoulo, A., Ramdani, B., Kawalek, P., & Lorenzo, O. (2009). Predicting SMEs' adoption of enterprise systems. *Journal of enterprise information management*.

Flynn, B. B., Huo, B., & Zhao, X. (2010). The impact of supply chain integration on performance: A contingency and configuration approach. *Journal of operations management, 28*(1), 58-71.

Gawankar, S. A., Gunasekaran, A., & Kamble, S. (2020). A study on investments in the big data-driven supply chain, performance measures and organisational performance in Indian retail 4.0 context. *International Journal of Production Research, 58*(5), 1574-1593.

Ghobakhloo, M., Arias-Aranda, D., & Benitez-Amado, J. (2011). Adoption of e-commerce applications in SMEs. *Industrial Management & Data Systems*.

Gillani, F., Chatha, K. A., Jajja, M. S. S., & Farooq, S. J. I. J. o. P. E. (2020). Implementation of digital manufacturing technologies: Antecedents and consequences. 107748.

Golgeci, I., & Ponomarov, S. Y. J. S. C. M. A. I. J. (2013). Does firm innovativeness enable effective responses to supply chain disruptions? An empirical study.

Gouda, S. K., & Saranga, H. (2018). Sustainable supply chains for supply chain sustainability: impact of sustainability efforts on supply chain risk. *International Journal of production research, 56*(17), 5820-5835.

Govindan, K., Cheng, T., Mishra, N., & Shukla, N. (2018). Big data analytics and application for logistics and supply chain management: Elsevier.

Govindan, K., & Fattahi, M. (2017). Investigating risk and robustness measures for supply chain network design under demand uncertainty: A case study of glass supply chain. *International Journal of Production Economics, 183*, 680-699.

Grover, P., & Kar, A. K. (2017). Big data analytics: A review on theoretical contributions and tools used in literature. *Global Journal of Flexible Systems Management, 18*(3), 203-229.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research, 70*, 308-317.

Gunasekaran, A., Subramanian, N., & Rahman, S. (2015). Supply chain resilience: role of complexities and strategies: Taylor & Francis.

Haimes, Y. Y., Matalas, N. C., Lambert, J. H., Jackson, B. A., & Fellows, J. F. J. J. o. i. s. (1998). Reducing vulnerability of water supply systems to attack. *4*(4), 164-177.

Haimes, Y. Y. J. R. A. A. I. J. (2009). On the definition of resilience in systems. *29*(4), 498-501.

Hendricks, K. B., & Singhal, V. R. (2005). An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production and Operations management, 14*(1), 35-52.

Hendricks, K. B., Singhal, V. R. J. P., & management, O. (2005). An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *14*(1), 35-52.

Hopkins, M. S., & Brokaw, L. J. M. S. M. R. (2011). Matchmaking with math: How analytics beats intuition to win customers. *52*(2), 35.

Hsu, P.-F., Ray, S., & Li-Hsieh, Y.-Y. (2014). Examining cloud computing adoption intention, pricing mechanism, and deployment model. *International Journal of Information Management, 34*(4), 474-488.

Ifinedo, P. (2011). Internet/e-business technologies acceptance in Canada's SMEs: an exploratory investigation. *Internet Research*.

Ismail, W., & Ali, A. (2013). Conceptual model for examining the factors that influence the likelihood of computerised accounting information system (CAIS) adoption among

malaysian SME. *International Journal of Information Technology and Business Management, 15*(1), 122-151.
Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review, 136*, 101922.
Kamalahmadi, M., & Parast, M. M. J. I. J. o. P. E. (2016). A review of the literature on the principles of enterprise and supply chain resilience: Major findings and directions for future research. *171*, 116-133.
Kamble, S. S., & Gunasekaran, A. (2020). Big data-driven supply chain performance measurement system: a review and framework for implementation. *International Journal of Production Research, 58*(1), 65-86.
Kauppi, K., Longoni, A., Caniato, F., & Kuula, M. (2016). Managing country disruption risks and improving operational performance: risk management along integrated supply chains. *International Journal of Production Economics, 182*, 484-495.
Kilubi, I., & Haasis, H. (2015). Supply chain risk management enablers-A framework development through systematic review of the literature from 2000 to 2015.
Kilubi, I., & Rogers, H. (2018). Bridging the gap between supply chain risk management and strategic technology partnering capabilities: insights from social capital theory. *Supply Chain Management: An International Journal, 23*(4), 278-292.
Kitano, H. J. N. R. G. (2004). Biological robustness. *5*(11), 826-837.
Klibi, W., & Martel, A. (2012). Scenario-based supply chain network risk modeling. *European Journal of Operational Research, 223*(3), 644-658.
Klibi, W., Martel, A., & Guitouni, A. (2010). The design of robust value-creating supply chain networks: a critical review. *European Journal of Operational Research, 203*(2), 283-293.
Krause, D. R., Vachon, S., & Klassen, R. D. J. J. o. S. C. M. (2009). Special topic forum on sustainable supply chain management: introduction and reflections on the role of purchasing management. *45*(4), 18-25.
Kshetri, N. J. I. J. o. I. M. (2018). 1 Blockchain's roles in meeting key supply chain management objectives. *39*, 80-89.
Kuan, K. K., & Chau, P. Y. (2001). A perception-based model for EDI adoption in small businesses using a technology–organization–environment framework. *Information & management, 38*(8), 507-521.
Kwak, D.-W., Seo, Y.-J., & Mason, R. (2018). Investigating the relationship between supply chain innovation, risk management capabilities and competitive advantage in global supply chains. *International Journal of Operations & Production Management, 38*(1), 2-21.
Kwon, O., Lee, N., & Shin, B. J. I. j. o. i. m. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *34*(3), 387-394.
Lai, K.-h., Wong, C. W., & Cheng, T. E. J. C. i. I. (2006). Institutional isomorphism and the adoption of information technology for supply chain management. *57*(1), 93-98.
Lai, Y., Sun, H., & Ren, J. (2018). Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management. *The International Journal of Logistics Management*.
LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review, 52*(2), 21-32.
Li, X. J. J. o. M. I. S. (2009). Managerial entrenchment with strategic information technology: A dynamic perspective. *25*(4), 183-204.
Lian, J.-W., Yen, D. C., & Wang, Y.-T. (2014). An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital. *International Journal of Information Management, 34*(1), 28-36.
Liang, H., Saraf, N., Hu, Q., & Xue, Y. J. M. q. (2007). Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. 59-87.
Lin, H. F., & Lee, G. G. (2005). Impact of organizational learning and knowledge management factors on e-business adoption. *Management Decision*.
Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world: Taylor & Francis.
Mackay, J., Munoz, A., & Pepper, M. J. J. o. R. R. (2019). Conceptualising redundancy and flexibility towards supply chain robustness and resilience. 1-21.

Maduku, D. K., Mpinganjira, M., & Duh, H. J. I. J. o. I. M. (2016). Understanding mobile marketing adoption intention by South African SMEs: A multi-perspective framework. *36*(5), 711-723.

Madzimure, J. J. E. J. o. B., & Management. (2020). Supply Chain Robustness, Firm Competitiveness and Performance of South African Logistics Firms. *8*(2), 137-149.

Maestrini, V., Luzzini, D., Maccarrone, P., & Caniato, F. J. I. J. o. P. E. (2017). Supply chain performance measurement systems: A systematic review and research agenda. *183*, 299-315.

Mahroof, K. J. I. J. o. I. M. (2019). A human-centric perspective exploring the readiness towards smart warehousing: The case of a large retail distribution warehouse. *45*, 176-190.

Manuj, I., Esper, T. L., & Stank, T. P. (2014). Supply chain risk management approaches under different conditions of risk. *Journal of Business Logistics, 35*(3), 241-258.

Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management strategies. *International Journal of Physical Distribution & Logistics Management, 38*(3), 192-223.

Marchand, D. A., Kettinger, W. J., & Rollins, J. D. J. M. S. M. R. (2000). Information orientation: people, technology and the bottom line. *41*(4), 69.

McAfee, A., & Brynjolfsson, E. (2008). Investing in the IT that makes a competitive difference. *Harvard business review, 86*(7/8), 98.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., & Barton, D. (2012). Big data: the management revolution. *Harvard business review, 90*(10), 60-68.

Mishra, N., & Singh, A. (2018). Use of twitter data for waste minimisation in beef supply chain. *Annals of Operations Research, 270*(1-2), 337-359.

Morash, E. A. J. T. j. (2001). Supply chain strategies, capabilities, and performance. 37-54.

Obal, M. J. I. M. M. (2017). What drives post-adoption usage? Investigating the negative and positive antecedents of disruptive technology continuous adoption intentions. *63*, 42-52.

Oliveira, T., & Martins, M. F. (2010). Understanding e-business adoption across industries in European countries. *Industrial Management & Data Systems*.

Oliveira, T., & Martins, M. F. (2011). Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation, 14*(1), 110.

Oliveira, T., Thomas, M., & Espadanal, M. (2014). Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management, 51*(5), 497-510.

Orji, I. J., Kusi-Sarpong, S., & Gupta, H. (2020). The critical success factors of using social media for supply chain social sustainability in the freight logistics industry. *International Journal of Production Research, 58*(5), 1522-1539.

Pettit, T. J., Croxton, K. L., & Fiksel, J. J. J. o. b. l. (2013). Ensuring supply chain resilience: development and implementation of an assessment tool. *34*(1), 46-76.

Ponomarov, S. Y., & Holcomb, M. C. J. T. i. j. o. l. m. (2009). Understanding the concept of supply chain resilience.

Pool, J. K., Arabzad, S. M., Asadi, A., Ansari, M. R. J. I. J. o. L. S., & Management. (2015). RFID acceptance in SMEs using TOE framework: an empirical investigation on Iranian SMEs. *21*(3), 335-347.

Raj, K. B., & Srinivas, K. J. I. J. S. C. M. V. (2019). Impact of robustness of supply chain on its performance-An empirical study. *8*(2), 290.

Ralston, P., & Blackhurst, J. J. I. J. o. P. R. (2020). Industry 4.0 and resilience in the supply chain: a driver of capability enhancement or capability loss? , 1-14.

Ranganathan, C., Teo, T. S., & Dhaliwal, J. (2011). Web-enabled supply chain management: Key antecedents and performance impacts. *International Journal of Information Management, 31*(6), 533-545.

Ransbotham, S., Kiron, D., & Prentice, P. K. (2015). The talent dividend. *MIT Sloan Management Review, 56*(4), 1.

Rao, S., & Goldsby, T. J. (2009). Supply chain risks: a review and typology. *The International Journal of Logistics Management, 20*(1), 97-123.

Riley, J. M., Klein, R., Miller, J., & Sridharan, V. (2016). How internal integration, information sharing, and training affect supply chain risk management capabilities. *International Journal of Physical Distribution & Logistics Management, 46*(10), 953-980.

Roßmann, B., Canzaniello, A., von der Gracht, H., Hartmann, E. J. T. F., & Change, S. (2018). The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study. *130*, 135-149.

Saenz, M. J., Koufteros, X., Durach, C. F., Wieland, A., Machuca, J. A. J. I. J. o. P. D., & Management, L. (2015). Antecedents and dimensions of supply chain robustness: a systematic literature review.

Sanders, N. R. (2014). *Big data driven supply chain management: A framework for implementing analytics and turning information into intelligence*: Pearson Education.

Shaik, M. N., Abdul-Kader, W. J. I. J. o. I. S., & Management, S. C. (2013). Interorganizational information systems adoption in supply chains: a context specific framework. *6*(1), 24-40.

Sila, I. J. E. c. r. (2013). Factors affecting the adoption of B2B e-commerce technologies. *13*(2), 199-236.

Srimarut, T., & Mekhum, W. J. I. J. S. C. M. V. (2020). Enhancing Financial Performance through Supply Chain Innovation (SCI) with Mediating Role of Robustness and Resilience Capability in Thailand's Manufacturing Sector. *9*(1), 175.

Srivastava, S. C., & Teo, T. S. (2010). E-government, e-business, and national economic performance. *Communications of the association for information systems, 26*(1), 14.

Stonebraker, P. W., Goldhar, J., & Nassos, G. J. J. o. M. T. M. (2009). Weak links in the supply chain: measuring fragility and sustainability.

Sugathan, S. K., Dominic, D. D., Ramayah, T., & Ratnam, K. A. (2015). The associating factors and outcomes of green supply chain management implementation–from the technological and non-technological perspectives *New Contributions in Information Systems and Technologies* (pp. 491-501): Springer.

Sun, S., Hall, D. J., & Cegielski, C. G. J. I. M. M. (2020). Organizational intention to adopt big data in the B2B context: An integrated view. *86*, 109-121.

Tan, K. H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics, 165*, 223-233.

Tang, C. S. (2006). Perspectives in supply chain risk management. *International Journal of Production Economics, 103*(2), 451-488.

Tarafdar, M., & Qrunfleh, S. J. I. J. o. P. R. (2017). Agile supply chain strategy and supply chain performance: complementary roles of supply chain practices and information systems capability for agility. *55*(4), 925-938.

Tarofder, A. K., Jawabri, A., Haque, A., Sherief, S. R. J. I. E., & Systems, M. (2019). Validating technology-organization-Environment (TOE) framework in web 2.0 adoption in supply chain management. *18*(3), 482-494.

Teece, D., & Pisano, G. (1998). The dynamic capabilities of firms: an introduction. In: Dosi, G., Teece, D. & Chytry, J., Technological, organization, and competitiviness–perspectives on industrial and corporate change: Nova York: Oxford University Press.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal, 18*(7), 509-533.

Teece, D. J. J. A. o. m. p. (2014). The foundations of enterprise performance: Dynamic and ordinary capabilities in an (economic) theory of firms. *28*(4), 328-352.

Teece, D. J. J. T. T., Know-How, L. o., & World, I. P. U. t. M. E. i. t. M. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *15*, 67-88.

To, M. L., Ngai, E. W. J. I. M., & Systems, D. (2006). Predicting the organisational adoption of B2C e-commerce: an empirical study.

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). *Processes of technological innovation*: Lexington books.

Tsou, C. W., Supakarapongkul, P., Pornjirattikal, S., & Huang, Y. T. (2014). *Research of Knowledge Transfer Mechanisms and Environmental Uncertaint towards Renewable Energy in Thailand.* Paper presented at the Applied Mechanics and Materials.

Urciuoli, L., & Hintsa, J. (2018). Improving supply chain risk management–can additional data help? *International Journal of Logistics Systems and Management, 30*(2), 195-224.

Ven, K., & Verelst, J. (2011). An empirical investigation into the assimilation of open source server software. *Communications of the association for information systems, 28*(1), 9.
Wagner, S. M., & Bode, C. (2008). An empirical examination of supply chain performance along several dimensions of risk. *Journal of business logistics, 29*(1), 307-325.
Wagner, S. M., Grosse-Ruyken, P. T., & Erhun, F. (2018). Determinants of sourcing flexibility and its impact on performance. *International Journal of Production Economics, 205*, 329-341.
Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics, 34*(2), 77-84.
Wamba, S. F., & Akter, S. (2019). Understanding supply chain analytics capabilities and agility for data-rich environments. *International Journal of Operations & Production Management*.
Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research, 70*, 356-365.
Wamba, S. F., Queiroz, M. M., & Trinchera, L. J. I. J. o. P. E. (2020). Dynamics between blockchain adoption determinants and supply chain performance: An empirical investigation. 107791.
Waters, D. (2011). *Supply chain risk management: vulnerability and resilience in logistics*: Kogan Page Publishers.
Wernerfelt, B. J. S. m. j. (1984). A resource-based view of the firm. *5*(2), 171-180.
Wieland, A., & Marcus Wallenburg, C. (2012). Dealing with supply chain risks: Linking risk management practices and strategies to performance. *International Journal of Physical Distribution & Logistics Management, 42*(10), 887-905.
Wong, L.-W., Leong, L.-Y., Hew, J.-J., Tan, G. W.-H., & Ooi, K.-B. (2020). Time to seize the digital evolution: Adoption of blockchain in operations and supply chain management among Malaysian SMEs. *International Journal of Information Management, 52*, 101997.
Yazdani, M., Zarate, P., Coulibaly, A., & Zavadskas, E. K. (2017). A group decision making support system in logistics and supply chain management. *Expert systems with Applications, 88*, 376-392.
Yu, W., Chavez, R., Jacobs, M. A., & Feng, M. (2018). Data-driven supply chain capabilities and performance: A resource-based view. *Transportation Research Part E: logistics and transportation review, 114*, 371-385.
Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q., Chen, X., & Zhang, T. (2015). A big data approach for logistics trajectory discovery from RFID-enabled production data. *International Journal of Production Economics, 165*, 260-272.
Zhou, Y., Sheu, J. B., & Wang, J. J. R. A. (2017). Robustness assessment of urban road network with consideration of multiple hazard events. *37*(8), 1477-1494.
Zhu, K., & Kraemer, K. L. (2005). Post-adoption variations in usage and value of e-business by organizations: cross-country evidence from the retail industry. *Information systems research, 16*(1), 61-84.
Zhu, K., Kraemer, K. L., & Xu, S. J. M. s. (2006). The process of innovation assimilation by firms in different countries: a technology diffusion perspective on e-business. *52*(10), 1557-1576.
Zhu, Y., Li, Y., Wang, W., & 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), 265-276.
Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization science, 13*(3), 339-351.

DECISION SCIENCES INSTITUTE

A Survey of the Three Main Categories of Business Analytics

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

ABSTRACT

In an increasingly competitive world, for both public and private sector organizations, survival may depend on how decision makers translate data analysis results into actionable and usable information for decision-making. Thus, decision makers across various industries employ three main categories of descriptive, predictive, and prescriptive analytics methods to make data-driven and evidence-based decisions. This study surveys the three main categories of business analytics and the specific methods within the categories. We are particularly interested in exploring the subject areas and domains in which those methods are used and the frequency with which the methods are employed in the published literature between the years 2010 and 2020.

KEYWORDS: Business Analytics, Descriptive, Predictive, and Prescriptive Analytics

INTRODUCTION

From a Business Analytics (BA) perspective, data is considered one of the most important assets and resources for companies operating in various industries. Therefore, to capitalize on data, decision makers are investing in new data analytics tools, methods, and techniques to gain insights into their operations, and guide their decision making. In other words, companies across various industries employ BA models, methods, applications to analyze both structured and unstructured data from a variety of sources to support and guide their data-driven business decision-making that are likely to impact the entire company.

It may be argued that BA is seen as indispensable to stay competitive (Appelbaum et al., 2017) as various research studies have discovered strong relationships between a company's performance in terms of profitability, revenue, and shareholder return, and its use of analytics (Evans, 2012). While the literature on BA offers ample evidence for BA's positive impact on organizational performance, the primary reason for most decision makers to employ BA in their organizations is that it improves organizational performance through enabling superior decision-making processes (Sharma et al., 2014).

In an increasingly competitive world, for both public and private sector organizations, survival may depend on how decision makers translate data analysis results into usable information for decision-making. Hence, since with the advent of better information gathering and processing tools, more structured insight-building from information is currently the standard expectation in various industries (Banerjee et al., 2013), decision makers may want to utilize various descriptive such as descriptive statistical measures, predictive such as regression analysis, and prescriptive such linear optimization BA methods to make evidence based and data driven strategic and tactical decisions.

This study surveys the three main categories of BA methods by focusing on how widely they are used in various publication outlets such as research articles and conference proceedings and the subject areas in which they are most commonly employed between the years 2010 and 2020.

LITERATURE REVIEW

While there is no agreed upon definition of BA, it may be defined as the process of using quantitative methods to derive meaning from data in order to make informed business decisions (Gavin, 2019), which could range from routine monitoring of business performance to more directed diagnosis of root cause of business problems as well as strategic prediction about future business initiatives (Banerjee et al., 2013).

Broadly speaking, BA can be broken down into three main categories that businesses routinely employ: descriptive, predictive, and prescriptive analytics (Davenport, 2015).

Most analytics start with descriptive analytics (Evans, 2012) which analyzes data from past occurrences and activities (Krishnan, 2020) to answer the question of “what happened and/or what is happening?” (Delen and Demirkan, 2013). Descriptive data analytics examines historical data to investigate how a business process, department, or product has performed over a particular time and to uncover details such as the frequency of events, the cost of operations, and the root cause of failures (IBM, 2013). Descriptive analytics techniques categorize, characterize, consolidate, and classify data to convert it into useful information for the purposes of understanding and analyzing business performance (Evans, 2012).

Predictive analytics uses a variety of models and techniques to predict future outcomes based on historical and current data (Gandomi and Haidar, 2015). It aims to observe and determine patterns in the dataset to predict future outcomes (Armoogum and Li, 2019) and to help in recognizing the trends and formative probabilities of uncertain outcomes (Bhattacharyya et al. 2020). The focus in predictive analytics is on the detection of repeated patterns of values in the data that can be used to make accurate predictions of future outcomes (Nisbet et al., 2018). It is what translates big data into meaningful, usable business information (Abbott, 2014). Predictive models and algorithms can detect relationships of any type between the predictor variables and the outcome variables and approximate them closely to make accurate predictions about future events (Nisbet et al., 2018). The ultimate purpose of any predictive analytics model is to convert data into knowledge and ultimately help make well-informed decisions (Kannabiran and Miller, 2012).

Prescriptive analytics may be defined as the application of testing and other techniques to determine which outcome will yield the best result in a given scenario (Gavin, 2019). Prescriptive analytics leverages data and mathematical algorithms for the prescription of specific actions (Delen & Demirkan, 2013). In other words, it seeks to find the best course of action for the future (Lepenioti et al., 2020). Prescriptive analytics models use optimization to identify the best alternatives to minimize or maximize some objective (Evans, 2012). It often utilizes simulation and optimization methodologies and techniques to generate a support mechanism, answering the question “What should we do” (Koukaras and Tjortjis, 2019).

BA allows companies to make sense of raw data they collect on internal processes and external business-related activities and helps transform raw data into actionable evidence-based insights. While companies utilize a variety of descriptive, predictive, and prescriptive BA tools and methods to analyze and transform data into useful and actionable information, the specific business problem the company is trying to solve may dictate which BA method will be employed.

RESEARCH METHOD

To explore and survey how frequently various BA methods within the three main categories of BA are employed in various publication outlets and to identify the subject areas in which the most commonly used BA methods are used, we used the data available on sciencedirect.com, which is one of the largest bibliographic databases of scientific publications maintained by Elsevier.

We first broke down BA methods into three main categories as suggested by the literature (Davenport, 2015): descriptive, predictive, and prescriptive analytics. Within each category we then identified some of the most commonly employed BA methods. For instance, with respect to prescriptive analytics, linear programming and Monte Carlo simulation are the two most frequently employed BA methods. Similarly, logistics regression and multiple regression methods are quite popular predictive analytics methods. Since, along with summary statistics, measures of association and cluster analyses may be used as descriptive statistics, both methods in this study are considered descriptive analytics methods (Table 1).

Having decided on which BA methods to survey and explore, we accessed the database and ran queries for each keyword tabulated in table 1. While a wealth of publication outlets is available on each BA method summarized in table 1, in this study, we focused on research articles, review articles, and conference abstracts published between the years 2010 and 2020, inclusive. Articles and conference proceedings containing any of the key words in its abstract or in the body of the publication was recorded and tallied in tables. Having retrieved the information on each method, we then refined our search by year, article type, and subject area.

Table 1: BA Categories and Methods

Descriptive Analytics	Predictive Analytics	Prescriptive Analytics
-Association Analysis -Cluster Analysis	-Multiple Regression -Logistic Regression -Neural Networks -Decision Trees	-Linear Programming -Monte Carlo Simulation

DATA ANALYSIS AND RESULTS

We first surveyed the terms “cluster analysis” and “association analysis;” the two most commonly used methods of descriptive analytics. As seen in figure 1, between the years 2010-2020, the term “cluster analysis” is employed a lot more frequently than the term “association analysis” across research articles, review articles, and conference abstracts. While back in 2010, the term “cluster analysis” was used 3173 times in the three main publication outlets, association analysis was cited only 721 times in the same outlets.

Business Analytics Categories

Descriptive Analytics

Figure 1: Frequency of descriptive analytics methods

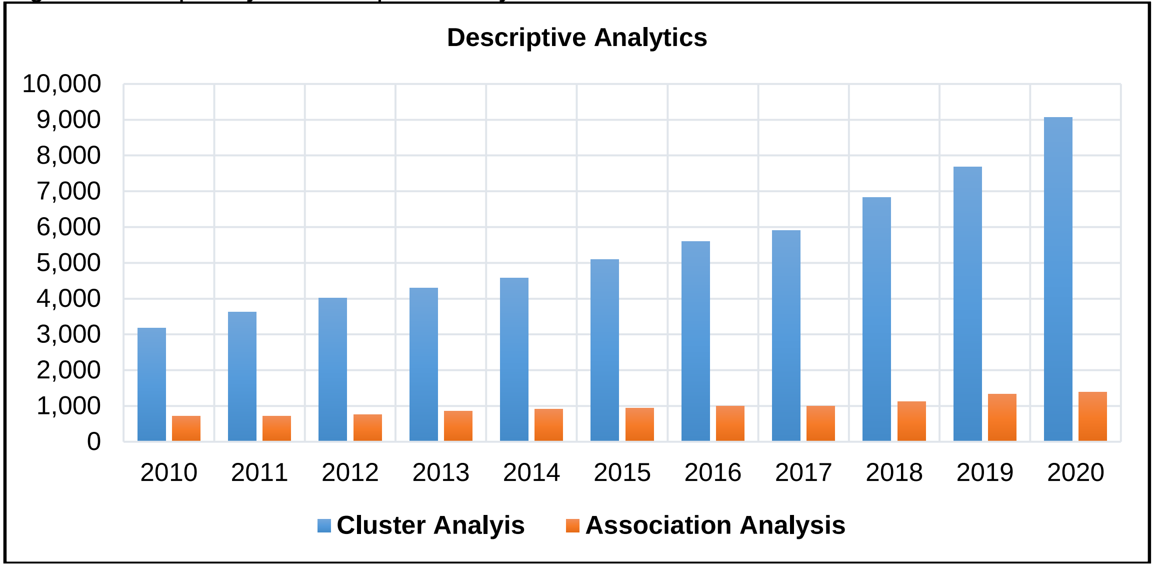


Table 2 summarizes the number of times both terms occur in the three main publication outlets between the years 2010 and 2020. As seen, cluster analysis is far more popular method of descriptive analytics compared to association analysis.

Table 2: Descriptive Analytics and Article Types

Method	Conference Abstracts	Research Articles	Review Articles	Total
Association Analysis	1184	8697	875	10756
Cluster Analysis	2380	54371	3106	59857
Grand Total	3564	63068	3981	70613

We also broke down descriptive analytics methods by the subject are in which they were employed and tabulated the results in table 3. As seen, both descriptive analytics methods are heavily employed in the Medicine and Dentistry domain, followed by Agricultural and Biological Sciences. Table 2 suggests that while association analysis is used in almost every subject area surveyed in this study, cluster analysis was not employed in such subject areas as Psychology, Business and Management, Computer Science, and Chemical engineering. Having said that, it should be pointed out that this study covers the subject areas between the years 2010 and 2020. Therefore, cluster analysis may have been used in the said subject areas in earlier studies.

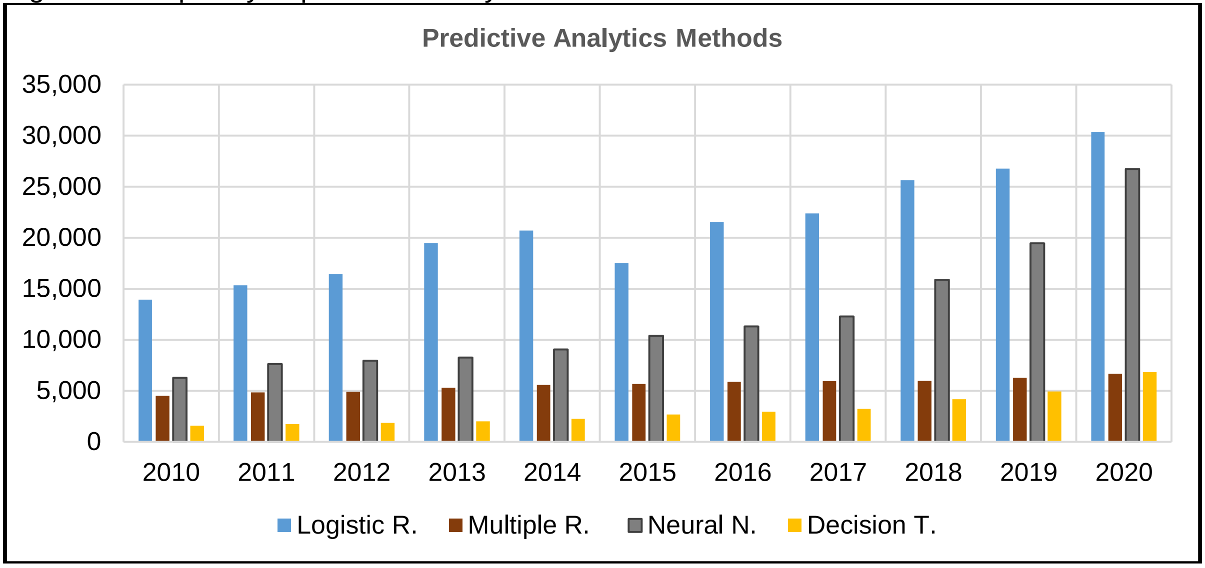
Table 3: Descriptive Analytics and Subject Areas		
Subject Area	Association Analysis	Cluster Analysis
Medicine and Dentistry	4995	11384
Agricultural and Biological Sciences	1459	13958
Biochemistry, Genetics and Molecular Biology	2926	10580
Environmental Science	474	11377
Earth and Planetary Sciences		7183
Neuroscience	1593	4465
Social Sciences	33	5388
Immunology and Microbiology	1004	3951
Chemistry	140	4013
Engineering	133	3902
Veterinary Science and Veterinary Medicine	739	
Psychology	543	
Pharmacology, Toxicology and Phar. Science	469	
Computer Science	395	
Decision Sciences	51	
Mathematics	42	
Business, Management and Accounting	37	
Chemical Engineering	32	

Predictive Analytics

With respect to predictive analytics, we surveyed four popular methods of logistic regression, multiple regression, neural networks, and decision trees. Figure 2 illustrates the number of times each term appears across different types of articles published between the years 2010 and 2020. As seen, among the four predictive analytics methods, logistic regression method seems to be the most commonly used predictive analytics method, followed by neural networks. Figure 2 suggests that decision trees, one of the most popular tools for classification and prediction, was employed a lot less frequently as a predictive analytics tool in the same time period.

Business Analytics Categories

Figure 2: Frequency of predictive analytics methods



.
Table 4 tabulates the number of times each key word appeared in the three main publication outlets between the years 2010 and 2020. It shows that research articles published between the years 2010 and 2020 alluded to the four predictive analytics methods a total of 375,741 times, followed by conference abstracts with a total of 63,816 times.

Table 4: Predictive Analytics And Article Types

Row Labels	Conference Abstracts	Research Articles	Review Articles
Logistic Regression	53785	170800	5443
Neural Networks	3073	120887	11200
Multiple Regression	4551	55075	1863
Decision Trees	2407	28979	2828
Grand Total	63816	375741	21334

Like in the case of descriptive analytics, Medicine and Dentistry subject area is one area in which all four methods are heavily used. Surprisingly, table 5 suggests that in the Business and Management subject area, decision trees, logistic regression, and neural networks methods are not mentioned between the years 2010 and 2020. Similarly, in the same years, the terms logistic regression, multiple regression, and neural networks do not appear in the Economics and Finance domain. Physics and Astronomy, Mathematics, and Materials Science subject areas seem to be relying heavily on neural network predictive analytics method as they mention no other predictive analytics methods in the same time period. While Business, Management and Accounting domain seems to have employed only multiple regression predictive analytics method, Economics, Econometrics and Finance, and Decision Sciences subject areas used only decision tree predictive analytics in the same time period.

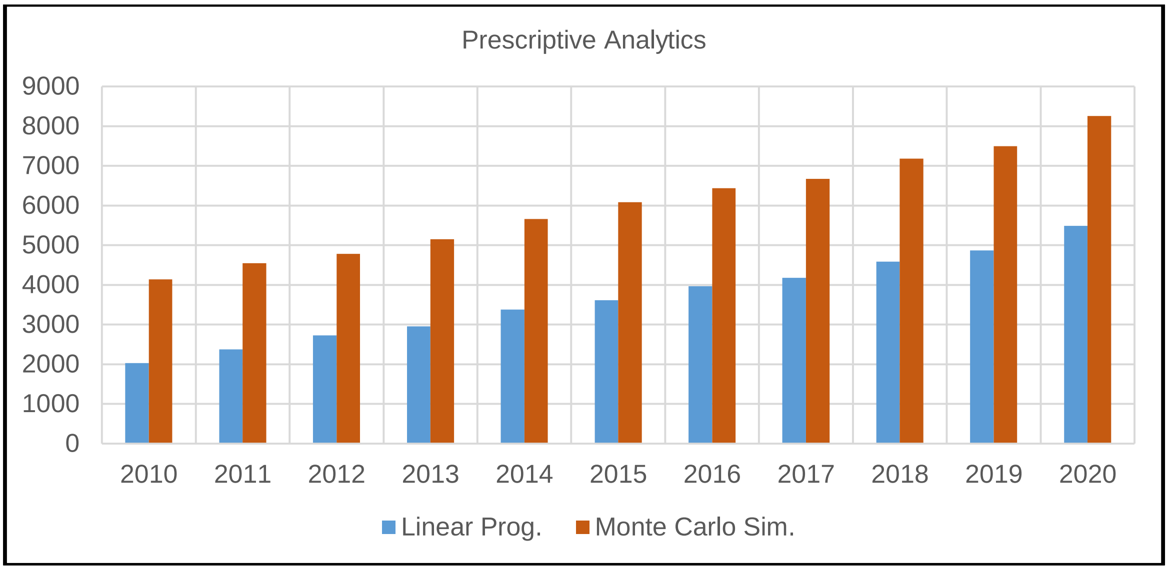
Table 5: Predictive Analytics and Subject Areas				
Subject Area	Decision Trees	Logistic R.	Multiple R.	Neural N.
Medicine and Dentistry	8590	182137	24287	15742
Engineering	7301		3832	46043
Computer Science	10580	6183		38196
Neuroscience		20534	9894	20943
Environmental Science	3978	8715	8112	9430
Social Sciences	3469	18152	7462	
Psychology		15331	9804	
Biochemistry, Genetics and Biology		18596	4878	
Agricultural and Biological Sciences	2763	10172	8346	
Energy	2047			15209
Earth and Planetary Sciences	2582		4046	7720
Immunology and Microbiology		11137		
Nursing and Health Professions		10999		
Physics and Astronomy				8473
Mathematics				7590
Materials Science				7417
Business, Management and Accounting			3037	
Economics, Econometrics and Finance	2004			
Decision Sciences	1743			

Prescriptive Analytics

In this section, we focus on the two most commonly employed prescriptive analytics methods; linear programming and Monte Carlo simulation. While Monte Carlo simulation is often used for modeling the probabilities of different outcomes, linear programming allows decision makers to achieve the most optimal solution for a problem with a set of constrains.

Figure 3 illustrates that the term “Monte Carlo simulation” is used a lot more frequently than linear programming in the three main types of articles published between the years 2010 and 2020, implying that Monte Carlo simulation is more popular prescriptive analytics method.

Figure 3: Frequency of prescriptive analytics methods



Similar statistics is presented in table 6, which tabulates the number of times both methods appeared in the three main publication outlets. As seen, across three publication categories, Monte Carlo simulation prescriptive analytics method seems to be more widely employed than linear programming.

Table 6: Prescriptive Analytics and Article Types

Method	Conference Abstracts	Research Articles	Review Articles
Linear Programming	114	38418	1628
Monte Carlo Simulation	2275	61007	3124
Grand Total	2389	99425	4752

With respect to the subject areas in which the two prescriptive analytics methods are employed, Engineering, Computer Science, and Energy domains are the top three subject areas where the two methods are heavily used when running prescriptive analytics. While Medicine and Dentistry domain was one particular domain where both descriptive and predictive analytics methods were commonly used, in terms of prescriptive analytics, it ranks a lot lower.

Subject areas such as Social Sciences, Chemical Engineering, and Business and Management seem to be using only linear programming prescriptive analytics methods. Similarly, subject areas such as Material Science, Medicine and Dentistry, and Earth Sciences appear to prefer Monte Carlo simulation method when seeking the most optimal solution to a problem.

Table 7: Prescriptive Analytics and Subject Areas		
Subject Area	Linear Programming	Monte Carlo Simulation
Engineering	16812	16662
Computer Science	10786	6456
Energy	9390	7630
Physics and Astronomy		15147
Mathematics	7700	6054
Environmental Science	3457	6656
Materials Science		7781
Decision Sciences	7689	
Economics, Econometrics and Finance	2024	5528
Medicine and Dentistry		7455
Earth and Planetary Sciences		5219
Social Sciences	2950	
Chemical Engineering	2206	
Business, Management and Accounting	1787	

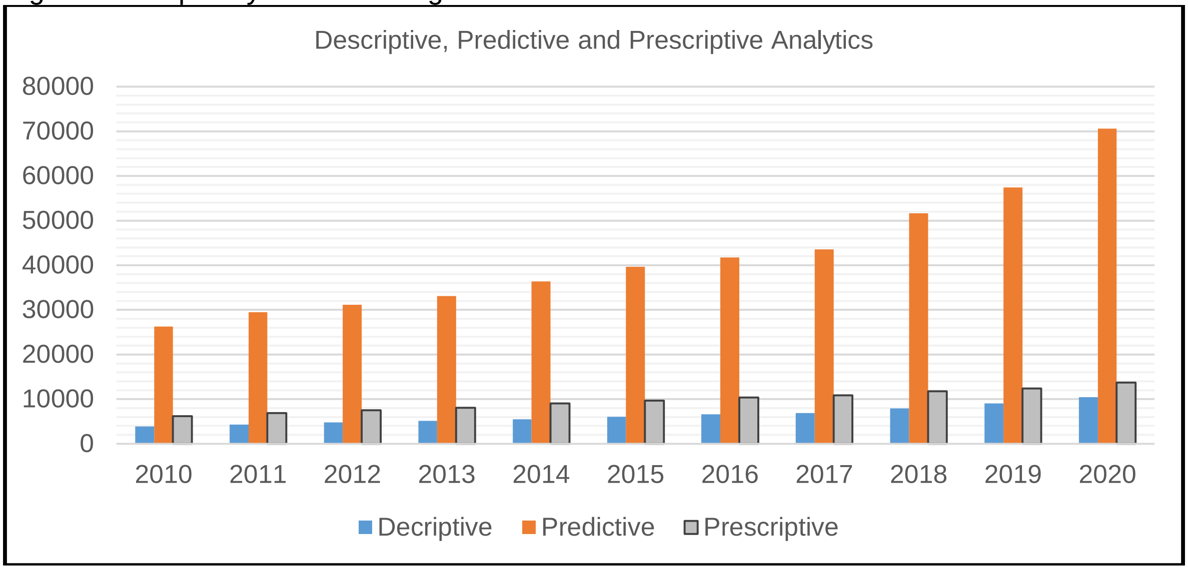
In addition to surveying each BA category separately, we also made a comparison among the three BA categories. As tabulated in table 8, across the three categories of BA, it seems predictive analytics methods are more commonly used than the other two categories. It may be due to the fact that the predictive analytics category covers four different methods, in contrast, the other two categories survey only two methods.

Table 8: Three BA Categories and Methods		
Category	Method	Sum of Frequency
Descriptive		70628
	Association Analysis	10764
	Cluster Analysis	59864
Predictive		460991
	Decision Trees	34222
	Logistic Regression	230103
	Multiple Regression	61504
	Neural Networks	135162
Prescriptive		106577
	Linear Programming	40163
	Monte Carlo Sim.	66414

Figure 4 illustrates the three main categories of BA and the total number of times they are alluded to between the years 2010 and 2020. Looking at figure 4, it seems predictive analytics methods are the most widely used data analytics methods compared to the other two categories.

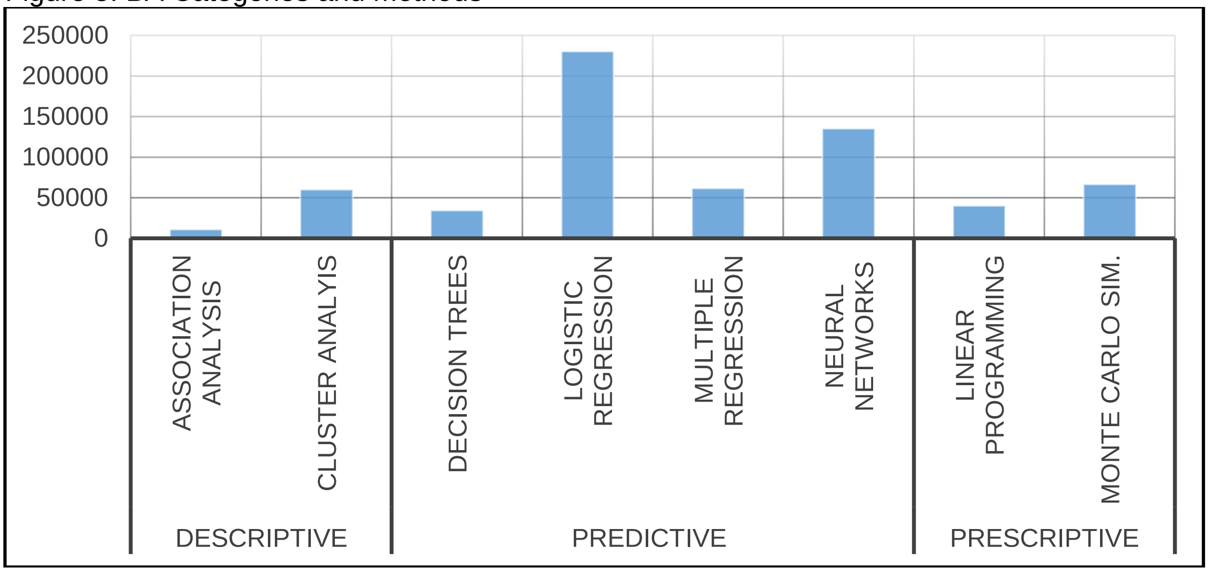
Business Analytics Categories

Figure 4: Frequency of three categories of BA between 2010 and 2020



Having drilled down into each category, it appears logistic regression method is the most commonly employed data analytics method, followed by neural networks, both of which are predictive analytics methods (figure 5). The term “association analysis” is the least frequently appearing term as an analytics method between 2010 and 2020.

Figure 5: BA Categories and methods



We finally examined the article types in which BA methods are employed. As summarized in table 9, research articles published between 2010 and 2020 allude to the BA methods more frequently than the other two types of articles.

Table 9: Article Types and BA Methods			
Method	Conference abstracts	Research articles	Review articles
Descriptive	**3564**	**63068**	**3981**
Association Analysis	1184	8697	875
Cluster Analysis	2380	54371	3106
Predictive	**63816**	**375741**	**21334**
Decision Trees	2407	28979	2828
Logistic Regression	53785	170800	5443
Multiple Regression	4551	55075	1863
Neural Networks	3073	120887	11200
Prescriptive	**2389**	**99425**	**4752**
Linear Programming	114	38418	1628
Monte Carlo Simulation	2275	61007	3124
Grand Total	**69769**	**538234**	**30067**

Table 10 tabulates and summarizes the frequency of each method by year. A close examination of the table shows that both logistic regression and neural networks analyses were employed far more frequently in the published literature in 2010 through 2020, implying that the research questions addressed in various subject areas in the same time period involved mostly a binary dependent variable.

Table 10: Frequency of BA Methods by Year								
	Descriptive		Predictive				Prescriptive	
Year	Cluster A.	Association A.	Logistic R.	Multiple R.	Neural N.	Decision T.	Linear Prog.	M. Carlo Sim.
2010	3,173	721	13,923	4,519	6,263	1,569	2,028	4,137
2011	3,621	723	15,320	4,846	7,618	1,718	2,371	4,547
2012	4,022	764	16,438	4,911	7,940	1,867	2,729	4,780
2013	4,295	863	19,485	5,297	8,254	2,018	2,953	5,152
2014	4,584	909	20,699	5,564	9,066	2,243	3,382	5,663
2015	5,091	940	17,532	5,676	10,381	2,683	3,610	6,085
2016	5,599	993	21,544	5,867	11,298	2,952	3,970	6,433
2017	5,900	1,003	22,372	5,935	12,293	3,218	4,182	6,674
2018	6,826	1,129	25,649	5,962	15,886	4,174	4,586	7,185
2019	7,688	1,332	26,758	6,263	19,437	4,947	4,868	7,499
2020	9,065	1,387	30,383	6,664	26,726	6,833	5,484	8,259
Total	**59,864**	**10,764**	**230,103**	**61,504**	**135,162**	**34,222**	**40,163**	**66,414**

DISCUSSION AND CONCLUSIONS

Since it allows decision makers and business professionals to make evidence based and data driven decisions, over the last decade, BA has become an integral part of how decisions are made and how businesses operate. Therefore, to extract value and make sense of data, business professionals use variety of BA methods. While a number of BA applications and methods are available to decision makers to capture, process, transform, and analyze data, the specific business problem decision makers are trying to solve may dictate which BA method will be employed.

In this study, we surveyed the three main categories of BA; descriptive, predictive, and prescriptive analytics. While all three are used in different subject areas and domains by decision makers, across the three categories, predictive models seem to be more popular than the other two. In addition, across all different methods within the three categories, logistic regression is the most commonly used BA method, followed by neural networks.

With respect to the subject areas in which these methods are employed, Medicine and Dentistry is the top subject area in both descriptive and predictive analytics categories, followed by Agricultural and Biological Sciences (descriptive analytics), Engineering (predictive analytics), and Engineering and Computer Science (prescriptive analytics).

BA, along with commonly used analytics methods, continue to evolve. As new data management and analytics tools come into the scene, decision makers will be able to continue to transform data into useful information faster, and ultimately make better data driven business decisions.

REFERENCES

Armoogum, S., and Li. (2019). Big data analytics and deep learning in bioinformatics with hadoop, in Deep Learning and Parallel Computing Environment for Bioengineering Systems, 2019.

Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting, *International Journal of Accounting Information Systems, 25*, 29-44.

Banerjee, A., Bandyopadhyay, T., and Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? *Vikalpa, 38*(4), October-December 2013, 1-11.

Bhattacharyya, S., Snasel, V., Gupta, D., and Khanna, A. (2020). *Hybrid computational intelligence challenges and applications,* Academic Press.

Davenport, H.T. (2015). 5 Essential principles for understanding analytics, Harvard Business Review, Retrieved on May 4, 2021, from https://hbr.org/2015/10/5-essential-principles-for-understanding-analytics

Delen, D. and Demirkan, H. (2013). Data, information and analytics as services, *Decision Support Systems, 55*(1), 359-363.

Evans, C. (2012). Business Analytics: The next frontier for decision sciences, *Decision Line, 43*(2), pp.4-6.

Gandomi, A., and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35, 137-144.

Gavin, M. (2019). Business analytics: What it is and why it's important, Retrieved on April 29, 2021, from https://online.hbs.edu/blog/post/importance-of-business-analytics.

IBM. (2013). Descriptive, predictive, prescriptive: Transforming asset and facilities management with analytics, Retrieved May 3, 2021, from https://elearn.ing.unipi.it/pluginfile.php/186198/mod_resource /content/2 /IBM- on-analytics.PDF

Kannabiran, D., and Miller, D., (2012). Practical text mining and statistical analysis for non-structured text data applications, Editors, Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., Nisbet, R. Academic Press, 2012.

Krishnan, K. (2020). *Building big data applications*. Academic Press. 1st edition.

Lepenioti, K., Bousdekis, A., Dimitris, A., and Gregoris, M. (2020). Prescriptive analytics: Literature review and research challenges, *International Journal of Information Management*, 50, 57-70.

Nisbet, R., Yale, K., and Miner, G. (2018). Predictive analytics for population health and care, in handbook of statistical analysis and data mining applications (Second Edition), 2018, 235-258

Sharma, R., Mithas, S. & Kankanhalli, A. Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organizations. *European Journal of Information Systems*, 23, 433–441.

DECISION SCIENCES INSTITUTE
Sentiment analysis on adoption of blockchain in different sectors in India

Shameem Shagirbasha
Great Lakes institute of Management
Email: shawaja@gmail.com

ABSTRACT

This paper showcases the potential application of blockchain in different sectors and customer acceptance towards the adoption of blockchain technologies and the study is conducted by performing text mining and Sentiment Analysis. Blockchain initiatives have been increasing rapidly and emerging countries like India have already started building their block chain ecosystems. This research aims to find out the sentiment on implementation of blockchain technology across the mainstream sectors like banking and finance, public sector, manufacturing and logistics and supply chain etc. The author has used sentiment analysis to understand the current scenario in the different sectors and how can we go about achieving higher penetration of blockchain technology and also to find out the challenges we currently face in implementing the same. This research is aided by mining sentiments from research/Industry reports concentrated to India.

Keywords: blockchain, sentiment analysis, sectors

INTRODUCTION

Blockchain Technology captivated attention from governments, Organization and businesses and Academic researchers, with its potential application across major industries, namely Finance, Insurance, Supply Chain, Health Care, Energy, Gaming etcetera. Many consumers are intrigued to know about the underlying technology of cryptocurrencies and peer to peer transaction offered by this Technology. Blockchain can be defined as a public ledger in which all the transactions processed are stored in chain of blocks. The chain of blocks formed as new blocks append to it. Blockchain technology has the key characteristics, such as decentralisation, immutability and auditability (Zheng .Z, 2018). key feature of this technology is that transactions are not recorded centrally instead, each party involved in the transaction maintains a local copy of the ledger. Hence, it potentially reduces the role of middleman and allows people to transfer of digital property or data in a safe, secure and incontrovertible manner. The ledger is a linked list of blocks, each comprising a set of transactions. Transactions are broadcasted and recorded by each participant in the blockchain network (García-Bañuelos .L, Dec 2016).

This paper measures the consumers sensitiveness on adopting Blockchain technology by capturing emotions from Industry reports—for possible inputs for Blockchain Literacy and Inclusion. Cryptocurrencies have emerged as the first generation of blockchain technology. Cryptocurrencies are digital currencies build over Blockchain Technology and are embedded with cryptographic techniques and peer-to-peer network. The first and most popular example of cryptocurrencies is Bitcoin (Alharby .M, 2017), With a uniquely designed data storage structure, transactions in Bitcoin network could happen without any third party and the core technology to build Bitcoin is Blockchain, which was first proposed in 2008 and implemented in 2009 (Zheng .Z X. .., 2017) since it allows payment to be finished without any bank or any intermediary, Blockchain can be used in various financial services such as digital assets, remittance and online payment (Peters G.W, 2015) Though Blockchain Technology had its popularity in Financial services, Cryptocurrency and payment services since its inception, Currently the same technology expanded its footprints in other sectors namely Healthcare, E-Vote, Supply chain and Agriculture . Few examples of Blockchain Technology, in ride-sharing companies, the drivers and riders can create the more user-driven and value-oriented marketplace with the help of the distributed ledger. In the education sector, the block chain solutions can manage the verification procedures and secure students' data. The sentiment analysis, Text Mining and Natural Language Processing is gaining importance in the field of research and academic study. In this Big Data revolution, the abundance of data generated in an unprecedented way and accessibility of those data is become easier and increasing surge in social media platforms for Sharing Opinions, Reviews, Micro-Blog which captured the attention of Organizations, Product Developers, Researchers and governments to gain valuable insights from opinions and texts (Saberi .B, 2017)

Application of Blockchain across different sectors

Adoption of Blockchain in Finance and Inclusion

Although the potential of Blockchain is widely claimed to be at par with early commercial Internet, banking firms needs to understand the key features of the technology and how it can be used to solve the current business issues. While Internet enables exchange of data, Blockchain can involve exchange of value. Banks need to seek opportunities, determine feasibility and impact, and develop proof of concepts. Blockchain will face a hurdle in widespread adoption by financial institutions if the government regulation status remains unsettled (Jani.S, 2018)

Finance is essentially about money, and much of the financial system can run more easily on the blockchain if fiat money (dollars, euros, and rupees) could be transacted directly on the chain. However, the questions around regulations and the policies will have to be resolved through focused discussions with competent regulatory authorities and incorporation of their thought-process. They concluded that regulators should engage, intervene at early stage and shape the innovation (Varma.J, 2019)

Blockchain can help achieve financial inclusion in India as well as any economy of the world. The biggest and obvious fact about Blockchain is that it's a digital technology. But if the willingness to adopt mobile applications by bank users be increased in India, a Blockchain based financial solution can be achieved. The nature of distributed ledger governed via consensus protocol removes the need of third parties to facilitate transactions and hence transaction costs are reduced. Another advantage of this technology is that it reduces the transaction time substantially from days (traditional banking) to just a few minutes. This further help in adoption of digital technology when immediate money transfers are needed. Apart from imitating current financial practices, Blockchain, an immutable ledger, serves as a reliable source of credit history.

All these attributes of Blockchain makes it an unrestricted financial solution since it is distributed across all its participants, be it urban or rural. Rural Indians will start buying overseas goods, exporting their produce overseas, or doing digital work with the ability to make and collect cross-country payments. Blockchain is going to bring a significant transformation in the Banking Sector. It has the potential to disrupt the standard business models and make the existing systems obsolete. A secured database of client information should be developed and shared by different banks which can help in reducing time, effort, and cost in interbank transactions. In a bid to evolve towards cashless society, this can be an appropriate time for initiating suitable efforts towards digitizing the Indian rupee through Blockchain technology. Fintech and startups should closely work with government agencies and regulators to make sure that the legal and regulatory framework supports the use of Blockchain applications. Industry needs should be accessed and customized Blockchain solutions should be developed to handle current inefficiencies and problems. Adoption of Blockchain has some challenges like security, privacy, and scalability which needs to be tackled to deem it commercially viable. Awareness of Blockchain should be spread through various trainings, workshops, and by incorporating it within the curriculum in educational institutions. Extensive Research and pilot projects must be undertaken to commercialize the Blockchain solutions at large scale. In the years to come, Blockchain will evolve as a disruptive force in transforming Indian banking sector by making banking transactions safer, faster, transparent, and cost effective. It can strongly be recommended that the present time is apt for adoption of Blockchain in India (Gupta.A, 2018).

Adoption of Blockchain in Supply Chain & Logistics

The line of various points involved in producing and delivering goods, from the procurement stage to the end customer is defined as Supply chain. It can consist of various stages and locations. Hence, it has become more difficult to trace events in the entire chain. Moreover, due to the lack of transparency in the supply chain, it has become hard for the buyers and customers to be sure of the true value of the products and services. There are also challenges related to illegal events whose accountability is hard to investigate due to which the world faces problems of counterfeiting, forced labor and poor conditions in factories.

Moreover, lack of transparency is not the only challenge for logistics. The huge amount of data associated with products or documentation can easily be lost across the entire supply chain network. Parties do not share information concerning the place of origin of an asset to determine quality. Organizations tend not share all relevant information with other participants. As blockchain ensures transparency and security, it can be a good solution for fixing supply chains. Recording the transfer of products on the digital ledger as transactions, allows to identify the main data which is relevant to manage the supply chain.

The key features of blockchain would be very useful for application in the supply chain are Public availability gives the opportunity to track products form the place of origin to the end customer. It gives an opportunity for the participation of all parties in the supply chain due its decentralized structure. Cryptography-based and immutable nature gives high assurance of security. Walmart, Everledger, Provenance have already started projects to implement blockchain technology and consequently make their supply chain more efficient. (Krystsina Sadouskaya, 2017)

Adoption of Blockchain in Property Rights / Land Registry System

Over the past couple of years, India has witnessed rapid economic growth, with the GDP growing 7.6% in the last fiscal year. The current government must capitalize on the current economic momentum and use this to accelerate its reform agenda. One of the areas requiring regulatory attention in India is the property market. India must establish a standardized property rights regime if it aims to be an economic powerhouse. To bolster current systems, a decentralized, open, and transparent method of record-keeping needs to be instituted. Also, it must be supplemented by a legal framework capable of guaranteeing and enforcing property rights. A possible solution to the current recordkeeping issues is blockchain technology. Particular attention is being paid to how blockchain can be used for registries. A blockchain is an instrument that ensures veracity, making it the perfect recording system for anything closely. (Meghna Bal, 2017).

Adoption of Blockchain in Health Care

Blockchain has also gathered interest as a platform to improve the authenticity and transparency of healthcare data through several use cases, from maintaining permissions in electronic health records (EHR) to streamlining claims processing.
Physician credentialing: With the use of Blockchain, a trustworthy network of verified credentials can be obtained. Many healthcare organizations are negatively affected by

physician-credentialing which can be resolved by using the distributed ledger to have a centralized database of employment history artefact of every physician. Organization can check the frequency of an individual's artefact been used previously before hiring him/her. This speeds up the credentialing process to a huge extent.
Payment contracting: Payment in healthcare is a laborious process. Blockchain brings value-based initiatives to the table which allows organizations to design contracts accordingly. Blockchain ledger continuously tracks data from trusted sources inside and outside the clinical setting and determines who owes what to whom and when it is due. Instead of providing rebates, charge backs and bundled payment programs, Blockchain ledger creates incentives for timely-delivery, behavior and reward results (Bass, J., 2019).

HYPOTHESES

H1: Perceived ease of use positively affects one's sentiments towards using Blockchain Technology
H2: Perceived usefulness positively affects one's sentiments towards using Blockchain Technology
H3: Perceived risk negatively affects one's sentiments towards using Blockchain Technology
H4: Transparency positively affects reputation of Blockchain Technology
H5: Lack of regulations positively affects perceived risk of Blockchain Technology
H6: Disintermediation positively affects perceived usefulness of Blockchain Technology
H7: High computational power negatively affects perceived usefulness of Blockchain Technology
H8: Counter party fraud positively affects perceived risk of Blockchain Technology
H9: Compatibility positively affects perceived ease of use of Blockchain Technology
H10: Complexity positively affects perceived ease of use of Blockchain Technology

METHOD

The author did *text analysis* to understand Blockchain's impact on Indian industries right now. Our sources of data would be the consulting reports of big 4 and other consulting firms, certain industry reports, etc.
Text analysis identifies important information within the text itself. That is how it is qualitative in nature unlike text analytics. While text analysis includes many methods and techniques under its umbrella, our focus is on *word frequency* and *sentiment analysis*.
Word frequency can be used to list the words/terms or ideas that appear most often in a given text. For example, this can be helpful for us in identifying the terms or phrases that consumers of Blockchain technology use most often. This is done using a *word cloud*, essentially. Since the researcher will segregate the data industry wise first, these words can lead us to the factors affecting blockchain' s adoption in each of them. After obtaining the most used words, the *correlation* can be found between each of them to understand if they occur in pairs or if any other prominent pattern exists. All these steps will lead us to topic analysis which will help us understand what a given text is talking about.
Hence, Text Analysis will cover:

- Word Frequency
- Correlation
- Topic Modeling
- Sentiment Analysis

Sentiment Analysis is the method of deciding 'computationally' whether a piece of writing is positive, negative or neutral. It is achieved by incorporating natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores within a sentence or phrase to persons, subjects, themes and categories. Sentiment Analysis is also known as opinion mining, which extracts a speaker's viewpoint or behavior. It is the automated process that uses machine learning to classify text-based subjective data.

Since it has many practical applications, Sentiment Analysis is currently a subject of great interest and growth. Because information on the Internet is constantly growing publicly and privately accessible, a large number of opinions-speaking texts are available in review sites, forums, blogs and social media. Companies use Sentiment Analysis to analyze data such as tweets, survey responses, and product reviews, gain key insights, and make decisions based on data. Analysis of sentiment helps large business data analysts gauge public opinion, perform complex market research, track brand and consumer credibility, and interpret customer experiences.

Typically, in addition to identifying the opinion, such programs derive attributes from the expression for example:

- Polarity: Whether the speaker expresses a positive or negative opinion
- Subject: the matter being debated
- Opinion holder: the individual or entity expressing the opinion

Taking the help of the above mentioned analyzed data and results, the difference in opinion (if any) can be understood and find out its root causes. It can give us a better idea in terms of where Blockchain currently stands in its technology acceptance lifecycle in both the markets which will ultimately help in drawing inferences as to whether Blockchain's implementation in Indian industries can be improved or not.

For sentiment analysis, our sources of data would be news articles and statements of important personnel, etc. along with the data used for text analysis.

The author conducted the analysis on R and the following is the workflow:

1) As mentioned above, the first step is to segregate the data from industry reports sector wise into different .csv files for each sector.
2) Second step is to convert the data into corpus, which is essentially a list of all textual data.
3) Next step is to clean the data before mining as the text has issues like junk URLs, abbreviations, numbers, explicit words, etc in raw form. Looking at the source of data, the need to assess how much cleaning required is understood. This activity is highly contextual to source and objective.

Breakdown of cleaning process:

- Bring the whole data to lower case.
- As URLs don't give any context, they were deleted.
- Next removed stop words like is, of, a, an, the, almost, and, as, at, just, most, its etc.
- Remove numbers, white spaces, punctuations, any reference to abbreviations.
- Remove certain and specific words as needed. This step is extremely crucial as well as iterative. Based on the understanding of the topics, the author was able to add specific words to consider for stop words. These are purely dependent on context and understandable after only after a few iterations.

4) Once the data is cleaned, it is ready for Maths- frequency, correlation, topic modelling, emotional variance analysis. But since these operations can't be done on corpus, a new object was created and called tdm (term document matrix)/dtm (document term matrix). Tdm is transpose of dtm. This dtm/tdm is fed to Maths.
5) To obtain Word cloud, the author set different cap or thresholds of frequency which is again contextual and codependent with the word removal process explained above.
6) Now basically correlation of one word with other words was found with the correlation strengths set as (say 0.2), where 1 corresponds to high correlation and 0 to no correlation. This process will give us the understanding of words which are used in pairs, if any, for example- blockchain and security.
7) Then the author proceeded with an unsupervised learning technique- Topic Modelling. There are many algorithms for the same like LDA, SLDA, LSA, etc. LDA was used.

Topic Modelling will detect word and phrase patterns within the text and clusters word groups with similar expressions.

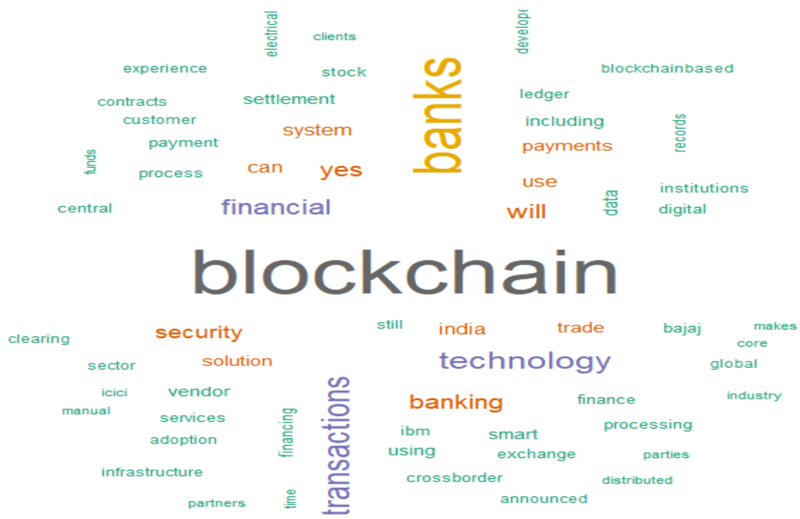
8) Lastly Sentiment Analysis was performed, also known as Emotional Variance Analysis. tdm/dtm is not required and it is done using *syuzhet* package. It detects about ten emotions- Positive, Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust and Negative.

RESULTS

Taking three sectors namely Public, Banking and Finance and Supply Chain and Logistics, Text Analysis was performed and following are the individual results obtained:
Each sector's result constitutes of its word cloud, few words with highest correlation and its explanation, Topic Modelling and lastly, sentiment analysis.

Banking And Finance

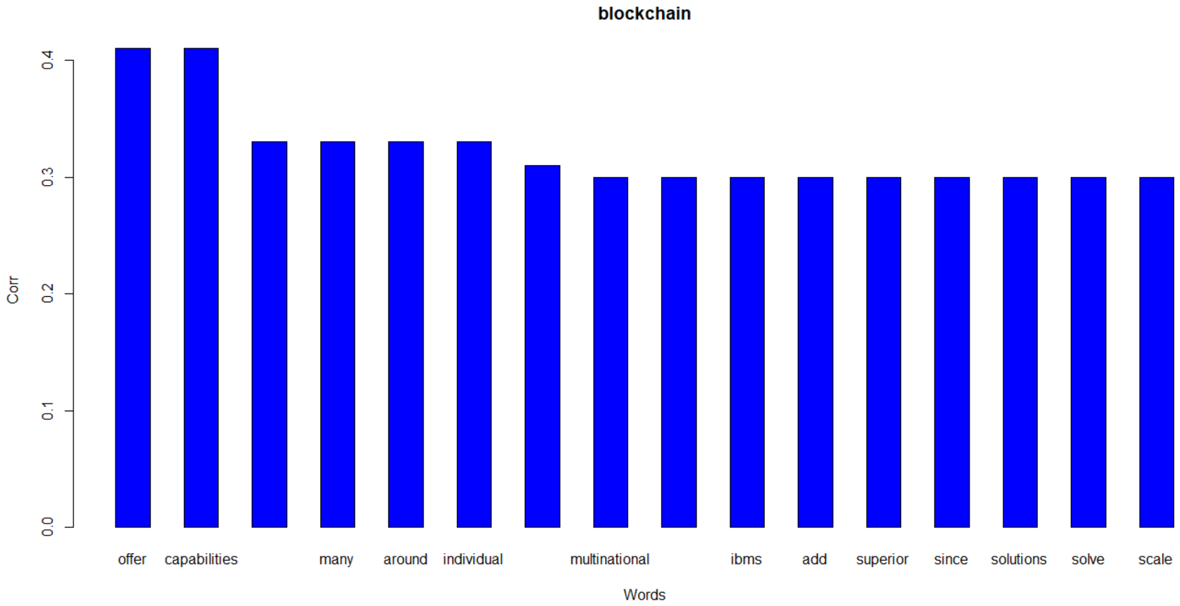
- **Word Cloud:**



- **Correlation:**

Blockchain

The word **blockchain** has high correlation with **offer** and **capabilities.**

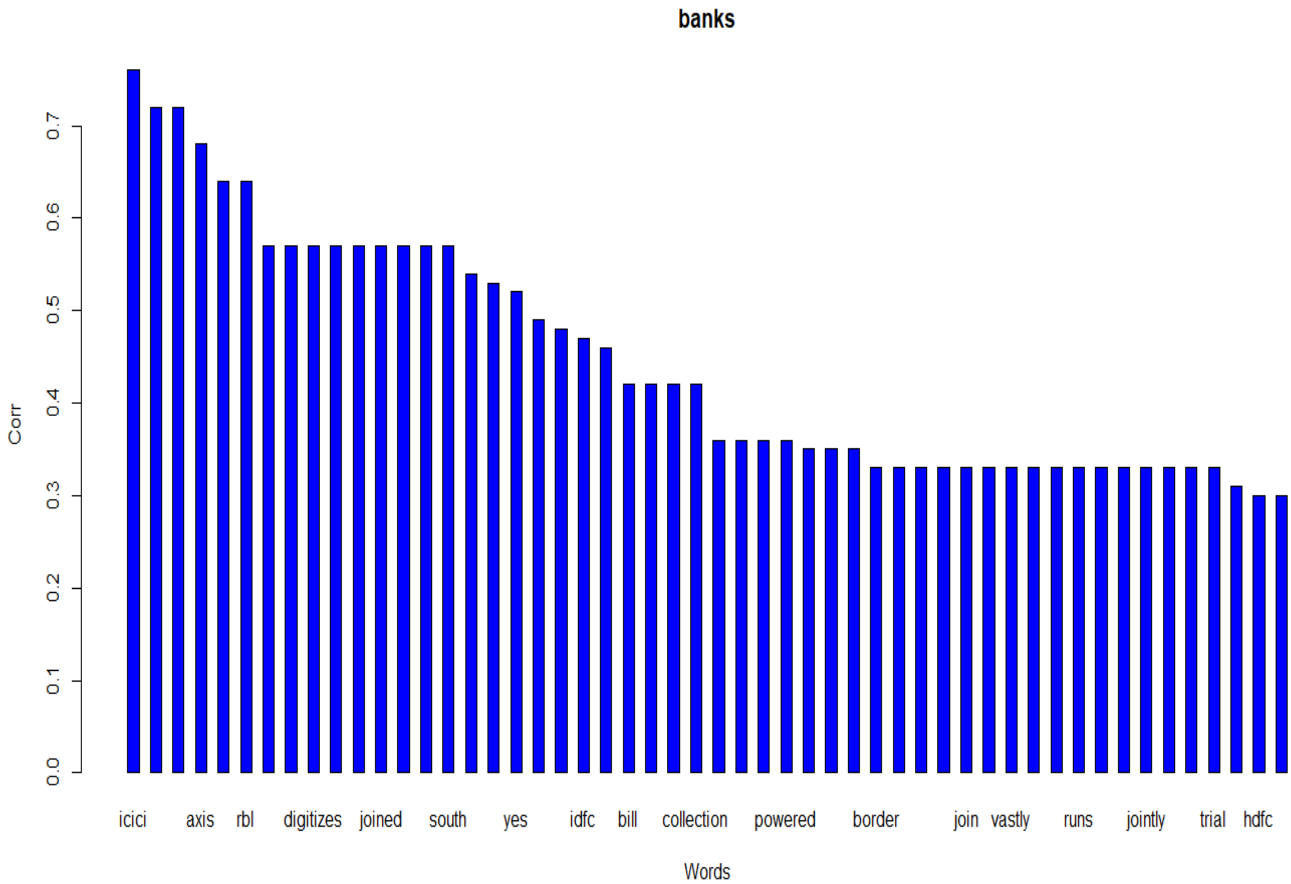


YES bank CEO and MD mentions using blockchain **capabilities** to solve the current challenges in Vendor Financing solutions.

Nasdaq was one of the first multinational financial services companies to begin using the blockchain **capabilities** in a non-currency manner.
Blockchain **offers** automated processing of transactions with almost zero manual intervention in Vendor financing by reducing process cycle for bill discounting, increased transparency, end-to-end digital process eliminating paper trail which is being used by YES bank with the help of **IBM** for its vendors Bajaj Electricals.

Blockchain based solution for trade finance requirements that digitizes trade finance business processes and **offers** functionality including bill collection, letters of credit and invoice financing was discussed in 'India Trade Connect' consortium led by Infosys which was joined by Axis Bank, ICICI Bank, IndusInd Bank, RBL Bank, Kotak Mahindra Bank, South Indian Bank, and Yes Bank.
Other words which are correlated with blockchain like **superior, many, solutions, solve** indicates its superior technology which useful in solving issues related to cross border payments, vendor financing etc.
The word **scale** indicates the difficulty in adoption of blockchain in a large scale due to high computational power required and hiring of data scientists or blockchain experts which is costly.

Banks



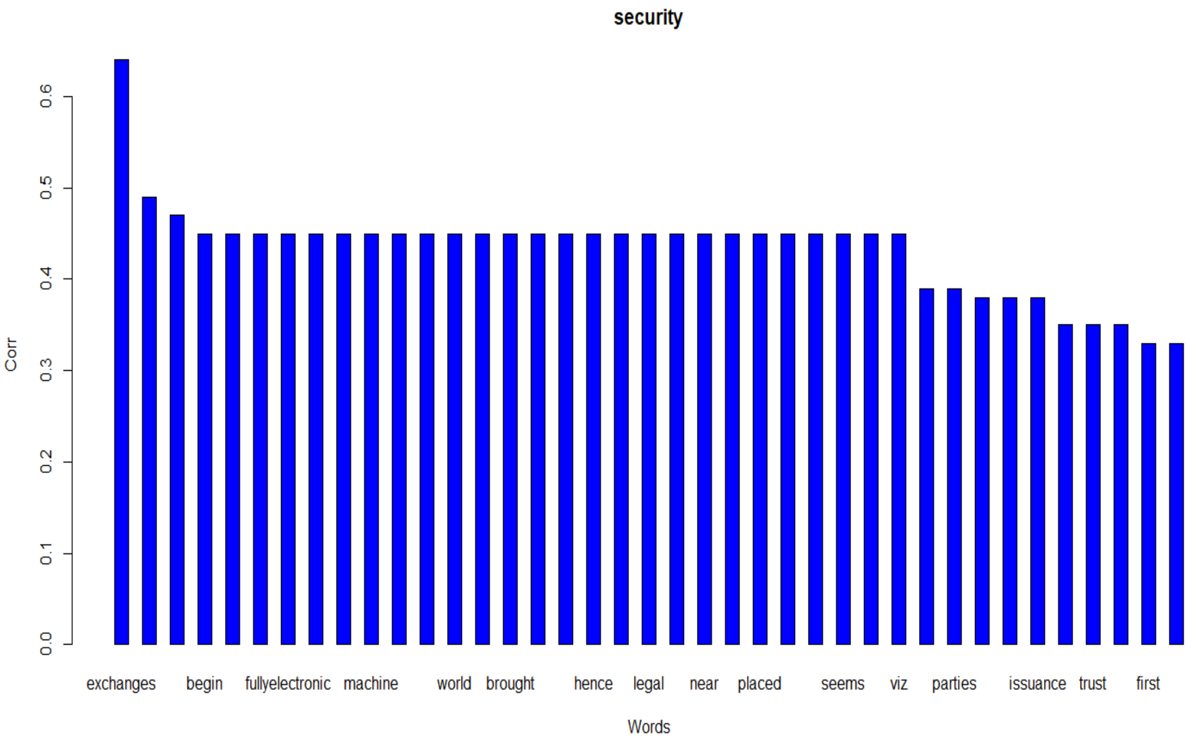
The above correlation with respect to **banks** tells us which companies in the banking and finance sector are implementing blockchain or exploring its capabilities for an efficient process. Banks like **Axis, HDFC, ICICI** are implementing blockchain for its KYC and trade finance procedures.

The seven banks including **ICICI Bank, Axis Bank, Federal Bank of India, Yes Bank** will collaborate with JP Morgan's Blockchain Interbank Information Network (IIN) which will help them improve the movement of funds and cross border payments.

Few banks like **Axis Bank, ICICI Bank, IDFC Bank, Kotak Mahindra Bank,** and **Yes Bank** have carried out pilot runs for vendor financing and international trade finance using Blockchain.

Border word indicates the steps that banks are taking to implement blockchain technology to address the challenges banks face in **cross-border payments**.

Security



Below are some of the negative views of the industry with regards to security (trust) of blockchain:

India, along with most of the countries, has its existing legal system set up with a **securities regulator** and **regulated exchanges** as the core parties. Hence, a new decentralized technology cannot by itself disrupt this legal model.
At least in the near term, it seems likely that **trust** will continue to be placed in central parties viz. the securities regulator and stock exchanges, unless **legal** amendments are brought in.

Below are some of the positive views of the industry with regards to security (trust) of blockchain:
Since on blockchain is immutable and can be used as a **trusted** third party for financial transactions.
Major fintech application where blockchain can be deployed are in Cross border payments which under the current system is time consuming , costly and requires a presence of a **trusted** third party. Blockchain can solve this since it can be used as a **trusted** third party.
The consensus system improves the transparency of decisions and **trust** among all stakeholders.

Topic Modelling:

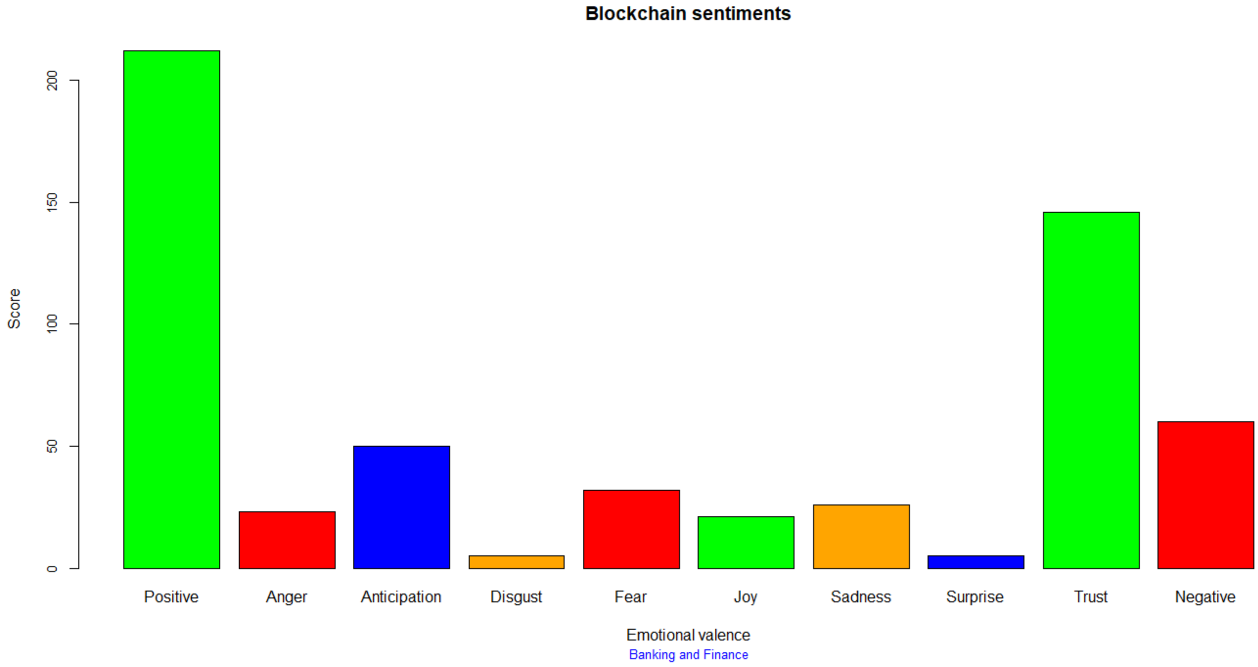
Topic	Keywords	Label
Topic 1	blockchain, banks, will, banking, technology, adoption, developers, partners, yes, ibm	Yes Bank and IBM, leading partners of blockchain adoption in banking
Topic 2	blockchain, financial, banks, trade, services, use, finance, digital, security, management	Financial trade security with blockchain
Topic 3	blockchain, transactions, will, payments, banks, can, crossborder, payment, smart, ledger	Crossborder payments with smart ledgers

Topic 4	blockchain, technology, financial, banking, system, infrastructure, institutions, processing, clearing, settlement	Institutions clearing financial settlements with blockchain technology
Topic 5	banks, blockchain, yes, exchange, transactions, vendor, stock, india, bajaj, financing	Indian banks exchanging transactions using blockchain.

Sentiment Analysis:
Overall, there is a **positive** sentiment around blockchains application in banking and finance sector with many companies who have either implemented or looking to implement it in their operations. Vendor financing, KYC, trade financing, cross border payments are some of the areas where the technology has been implemented. ICICI Bank, Axis Bank, Federal Bank of India, Yes Bank, JP Morgan, HDFC, IndusInd Bank, RBL Bank, Kotak Mahindra Bank and South Indian Bank are some of the companies having a understood the capabilities of the technology and its uses in their operations, have positive sentiments over its application.

The immutable nature of blockchain makes it a trusted third party hence in areas like cross border payments where you need a trusted third party to carry out the operations blockchain can be solution. Plus, the industry believes the consensus system improves the transparency of decisions and **trust** among all stakeholders. This is in line with study (Atzori, M. ,2017).

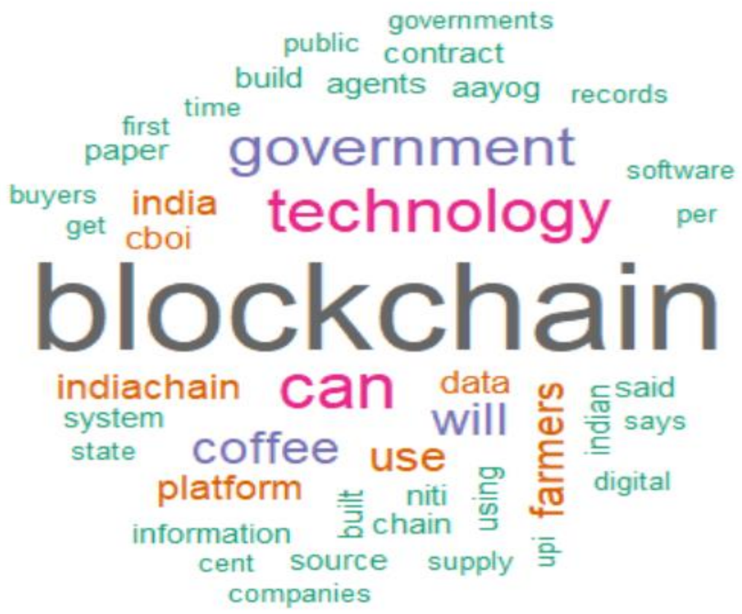
Lack of governance, non-scalability, and regulatory compliance risks are three of the major blockchain limitations for this sector. Possible misaligned motives of the members participating in a blockchain-powered financial transaction, high computational power required to process



the transactions and lack of regulations around using cryptocurrencies or smart contracts for transfers are some of the reasons for **negative** sentiments. This is in line with study (Dirk A. Zetzsche , Ross P. Buckley and Douglas W. Arner, 2017)

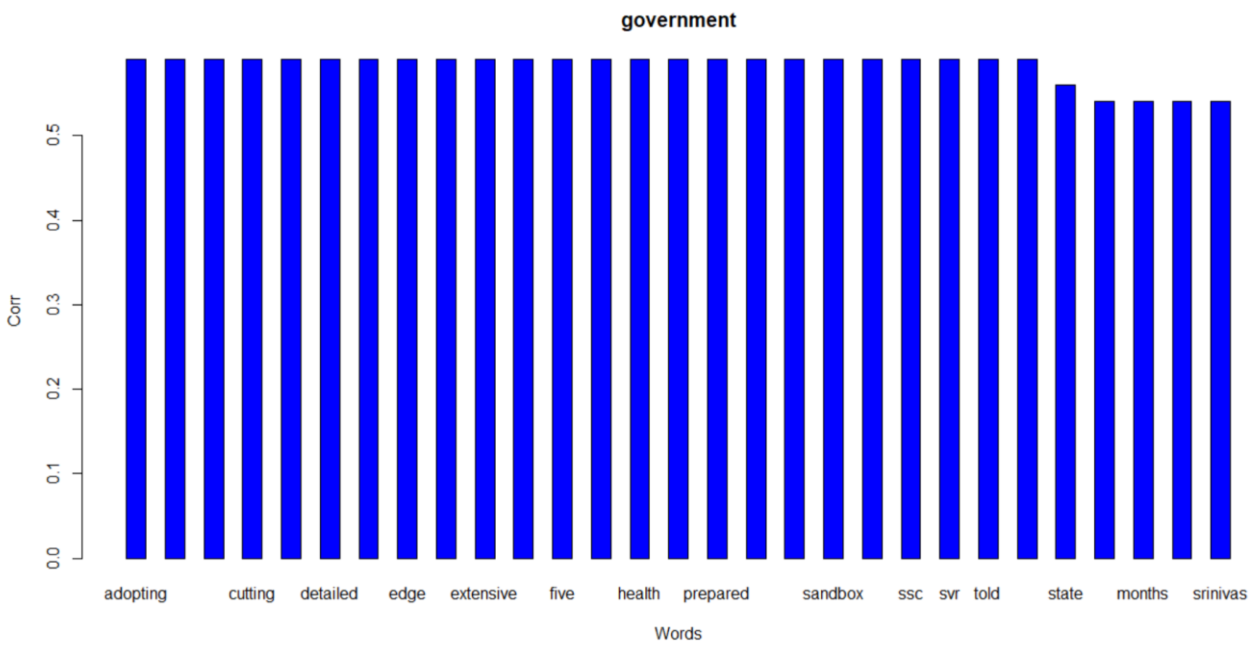
<u>Public sector</u>

- **Word Cloud**



- **Correlation**

<u>Government</u>



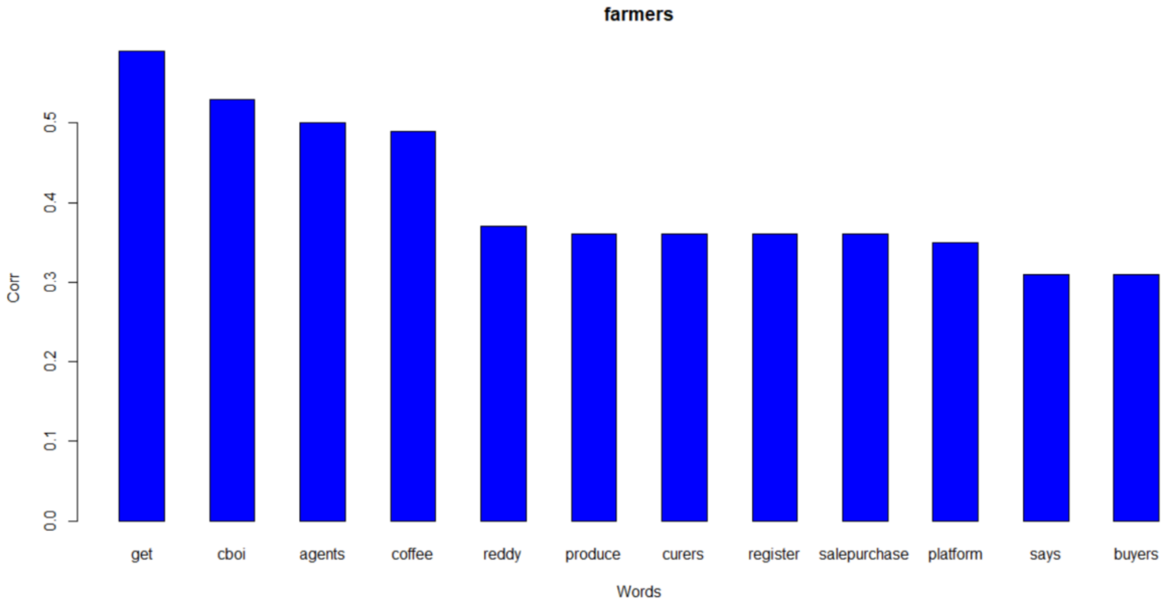
The word **government** has high correlation with many words like **adopting, cutting, edge, enhance, extensive** and **departments.**

Government's project IndiaChain has the capability of transforming **record-keeping** mechanisms, distribution of subsidiaries and tax monitoring. Nearly half of Indian states are involved in new Blockchain-based initiatives and states like Telangana and Goa co-hosted International Blockchain congress 2018 with NITI Aayog, aiming to bring Blockchain revolution in the country. Maharashtra is bidding on Blockchain to **enhance** service delivery.

Government has allocated Rs 4 crores of funds in 2019-20 for the **adoption** of this **cutting-edge** technology and it is believed that this **platform** will allow both government and private entities to bring scalable solutions. Officials said Blockchain is unhackable, leakage-proof and secure.

Information Technology department says the government has already launched pilot projects in the fields of health, documents, supply chain and SCC certificates and is coming forward with report on **extensive** use of Blockchain in other **departments**.

Farmers



The above correlation with respect to **farmers** tells us about Coffee Board of India (CBoI)'s venture of blockchain-based electronic-market platform for transparent and efficient sale of coffee produced by Indian farmers.

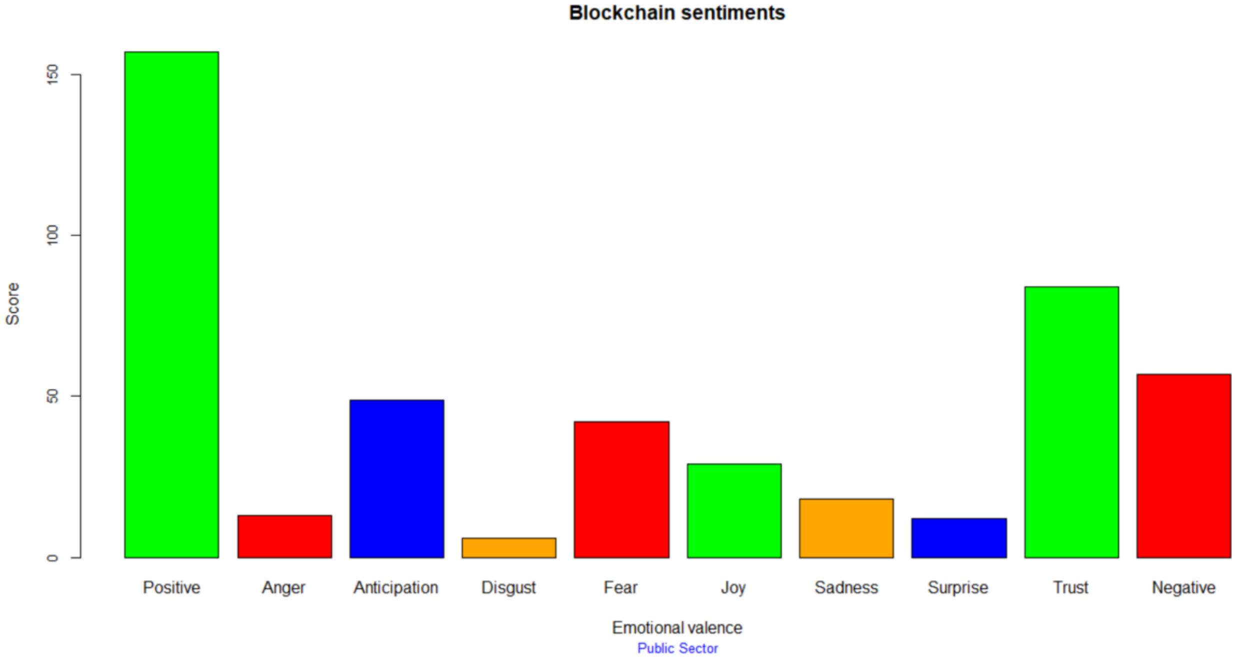
Initial response to this blockchain project by CBoI was very dull as **building trust** among farmers was a major issue but within months it gained momentum with huge no. of farmers, exporters, curers, roasters and international buyers registering on it. Another concern was the usage of cryptocurrencies for financial transactions which CBoI refrained from, as it is illegal in India.

Not only it provides farmers a better price of their produce but completely eliminates the use of **intermediaries**, i.e., the traditional channel of selling through **agents**. Etherium based blockchain application called **smart contract** connects farmers directly to buyers and allow them to draw sale-purchase contracts.

Topic Modelling

Topic	Keywords	Label
Topic 1	technology, will, coffee, blockchain, farmers, government, get, state, indian, cboi	Indian government aiding coffee farmers via blockchain
Topic 2	can, blockchain, coffee, platform, indiachain, blocks, sidechains, use, data, build	Sidechains on Blockchain platform
Topic 3	blockchain, can, use, technology, contract, government, agents, chain, supply, agriculture	Government using blockchain technology for agriculture supply chain
Topic 4	blockchain, technology, government, built, using, project, public, india, use, transactions	Indian government public projects on blockchain
Topic 5	blockchain, will, paper, india, aayog, niti, government, indiachain, indian, platform	Indiachain, a Niti Aayog blockchain platform

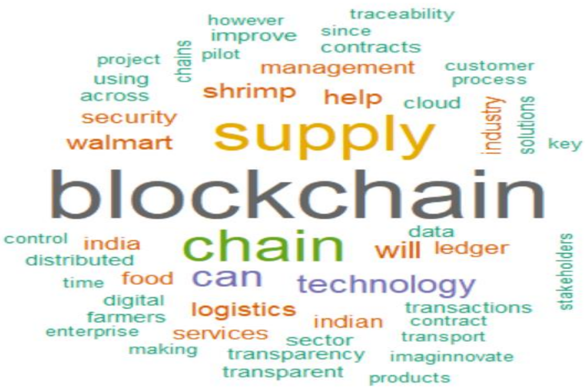
Sentiment Analysis



There is a **positive** sentiment and good amount of **anticipation** around blockchain-based applications in public sector with government and many states who have either implemented or looking to implement it in their operations. Blockchain can bring transparency and efficiency in public administration with government planning to step into the education sector to curb fraudulent degrees and certifications, and healthcare sector to manage huge amount of data. This is in line with study (Gartner Ray Valdes, David Furlonger and Fabio Chesini, 2016).
State governments are ensuring progressive approach to bring out a framework to implement this cutting-edge technology, while helping startups to participate in these initiatives.
The elimination of intermediaries like agents, in blockchain, enables farmers to get a better price for their produce, and establish direct relationship with buyers via smart contracts, though the use of crypocurrencies is not yet enabled. This is in line with study (Abhishek Gupta, 2018).
Due to its immutable nature, the government called in trustworthy, secure and unhackable but found it challenging to develop trust initially among its users, like farmers, who are used to trade traditionally which might be one of the reasons for **negative** sentiments.

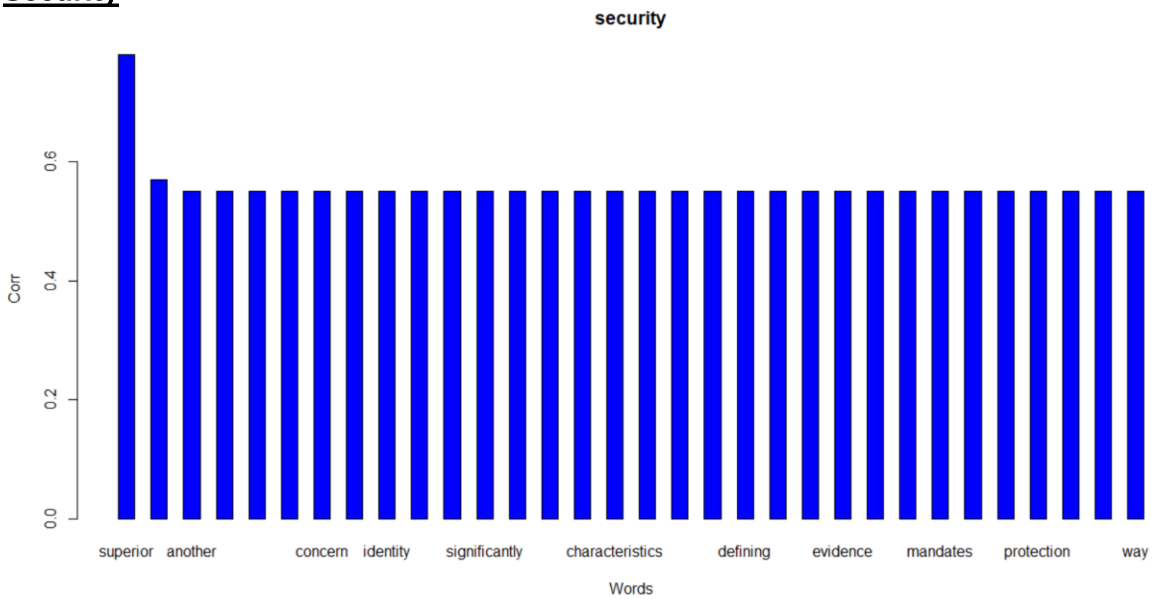
<u>Supply Chain and Logistics</u>

- **Word Cloud**



- **Correlation**

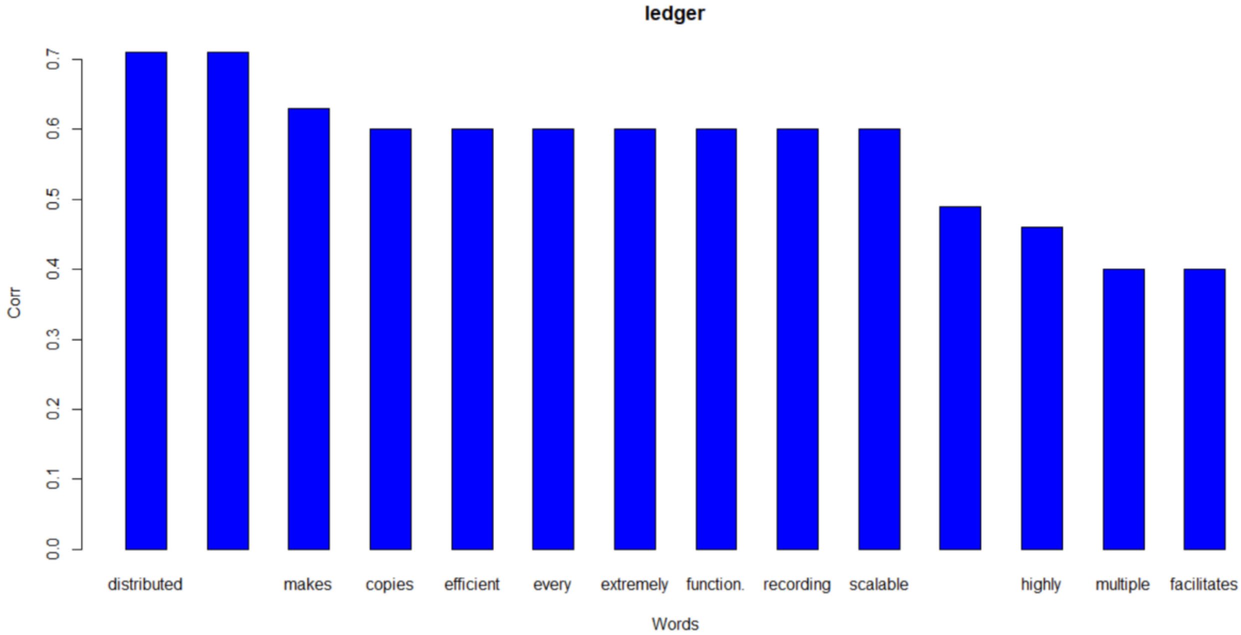
Security



The word **security** has high correlation with many words like **superior, blockchain-enabled, automating, infrastructure, strengthen, significantly, enabling** and many more along the same lines.

Blockchain, the digital ledger technology is critical in the growth of logistics sector as it protects data and increases supply chain transparency. Aiming at the security during third party supplier interactions, a Bangalore based startup called Signzy **enables** digital drafting and signing of legal contracts using Blockchain. FreightCrate Technologies also believes that Blockchain enables safe, cheap and effective communication. Through decentralization, encryption methods and immutable record keeping, a large amount of data can be shown to all the stakeholders without compromising the data security. It further enhances the security and governance by allowing smart contracts over cloud and their **automatic** management. Cloud services also require identity and access management which is also **strengthened** with **superior** levels of security and validation through blockchain-enabled digital identity using public key **infrastructure**. It seems, the combination of Blockchain and cloud is working well for many companies and helping stakeholders collaborate in a secure environment.

Ledger

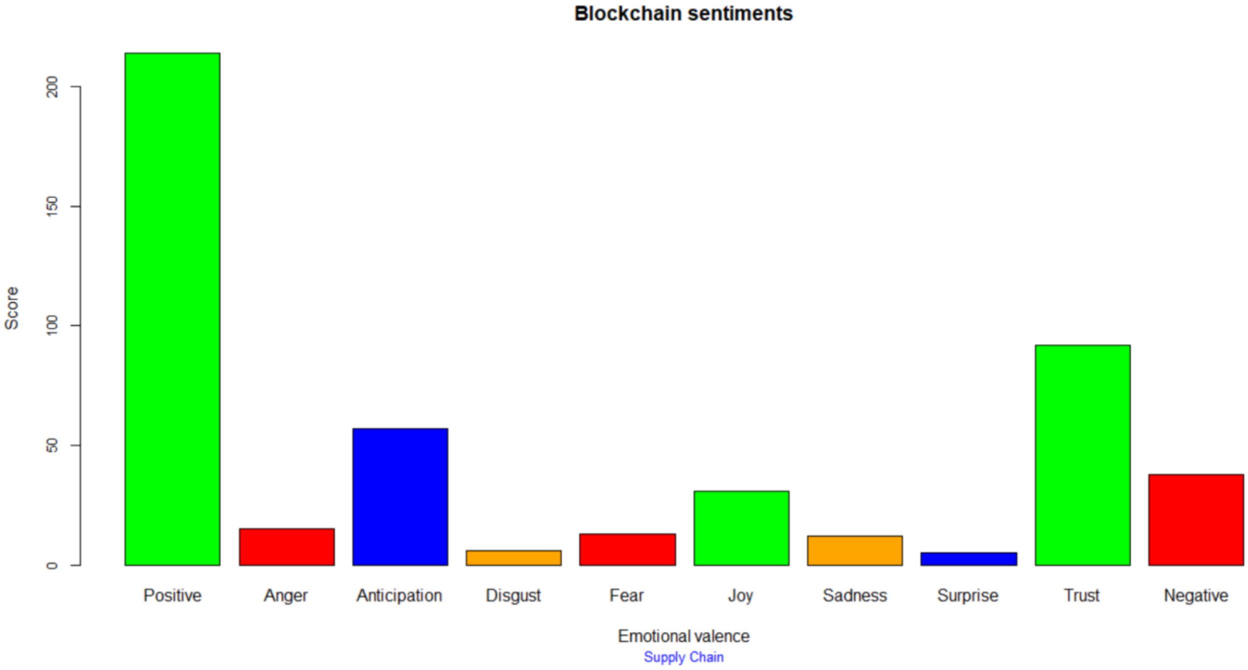


The word **ledger** has high correlation with many words like **distributed, transactions, efficient, extremely, scalable and transparent.**
Blockchain's biggest advantage seen by NITI Aayog is that it will be tamper-proof and transparent ledger that can help in fighting corruption at various levels of supply chain in businesses of India. In partnership with PwC and Intel, NITI Aayog started a pilot Blockchain application to track and streamline the fertilizer subsidy supply chain. All the supporting documents such as challans, invoices and claims are integrated in this application. A similar agricultural application by the name of eNAM is created to remove mistrust between farmers and intermediaries. Drugs are being pushed into the value chain by recording and time-stamping each transaction in the ledger to ensure safety and security. An Indian startup, OpenXcell helps its clients build customized supply chain according to the business goals which helps cater to the need of transparent multi-party transactions. India is also prone to piracy and sale of rip-offs along with scams, fraud and inefficient paper-based transactions which could change after the **effective** deployment of **distributed** ledger technology (DLT) in the sector. Not just domestic but business done with international parties can also be simplified and traceability be increased as professionals can track merchandise in the movement history. Financial issues can be resolved easily as **transactions** are **transparent**.

Topic Modelling

Topic	**Keywords**	**Label**
Topic 1	supply, chain, blockchain, ledger, across, highly, transactions, transparent, sector, time	Creating supply chain transparency with blockchain ledgers
Topic 2	blockchain, supply, chain, technology, help, can, ledger, distributed, since, will	Distributed ledger booming supply chain
Topic 3	blockchain, supply, chain, can, food, walmart, management, delivery, across, data	Blockchain data helping Walmart's food delivery chain
Topic 4	blockchain, will, supply, imaginnovate, transportation, industry, stakeholders, can, chain, logistics	Innovating transportation industry with logistics on blockchain
Topic 5	blockchain, services, can, cloud, shrimp, management, technology, sector, india, key	Cloud management services with blockchain technology

- **Sentiment Analysis**



There is a good amount of **positive** sentiment, **trust** and **anticipation** around Blockchain in Supply Chain and Logistics sector as it is one of the most obvious sectors to be benefitted from Blockchain's implementation.
Companies like Primechain bring in huge anticipation with the belief that blockchain technology will be an enabler of "massive social upliftment and economic prosperity" in India and with its several projects, including BankChain, Primechain-MONEY and a number of educational programs aimed at increasing blockchain adoption on a global scale. Imaginnovate, specialised in delivering technology solutions to transportation and logistics companies, has joined Blockchain in Transport Alliance (BiTA), where both the parties will be benefitted from each other's experience and knowledge, as they believe there is huge scope of improvement in terms of transport in logisctics. Walmart plans to use a blockchain food supply chain solution for shrimp farming in India to improve the supply chain traceability. All of these applications aim to bring data **security** to enable **tamper-proof record keeping.** This is in line with study (*A. Dorri, M. Steger, S. S. Kanhere and R. Jurdak,2017)*
Blockchain's immutable nature and it being a distributed ledger providing transparency in transactions and as a result, **traceability**, are the prime factors of the above companies' trust and anticipation. Removal of intermediaries, paper-work and ineffective communication is what is expected of this digital solution the most. This is in line with study (*Meghna Bal, 2017, Bass, J., 2019*).

DISCUSSION

This study provides evidence that perceived usefulness, perceived ease of use, perceived risk and reputation are related to blockchain implementation sentiments (accepting hypothesis H1, H2 and H3) which is in line with studies (Florian O. Knauer and Andreas Mann, 2019; C. Christopher Lee, John C. Kriscenski , Hyoun Sook Lim, 2019).The results indicate that transparency positively influences reputation of blockchain technology (accepting hypothesis H4). This is line with study (Andreas Kamilaris, Agusti Fonts and Francesc X. Prenafeta, 2020). Lack of regulation was found to have significant influence on perceived risk (accepting hypothesis H5) which is in line with study (Guych Nuryyev , Yu-Ping Wang , Jennet Achyldurdyyeva , Bih-Shiaw Jaw , Yi-Shien Yeh , Hsien-Tang Lin and Li-Fan Wu, 2020). The results further show that disintermediation positively influences perceived usefulness (accepting hypothesis H6) and high computational power negatively influences perceived usefulness (accepting hypothesis H7). This finding is in line with study (C. Christopher Lee, John C. Kriscenski, Hyoun Sook Lim, 2019).

Counter party fraud was found to have positive influence on perceived risk (accepting hypothesis H8). This is in line with study (Joe Abou Jaoude, Raafat Saade, 2007). However, the effect of compatibility on perceived ease of use was found to be insignificant (rejecting hypothesis 9). It was further found that the effect of complexity on perceived ease of use was insignificant (rejecting hypothesis 10). These findings deviate from previous study (Florian O. Knauer and Andreas Mann, 2019).

Managerial Implications:

There are significant managerial implications and insights from the findings of our study to better cope and coordinate the successful implementation of blockchain technology. The study helps identify important constructs for successful implementation of blockchain in Banking and Finance , Public and supply chain and logistics sectors and how the practitioner’s sentiments on adopting blockchain technology is influenced. The study also reveals that the adoption of blockchain mainly operates through perceived usefulness with many of the constructs directly effecting it. Hence for the companies implementing Blockchain technology, focus should be more on making blockchain technology more user-friendly and at the same time offering great value. The findings also reveal that reputation of the technology as being transparent influences its adoption. Based on how other practitioners have reacted after using the technology plays an important role in keeping a good reputation of the technology. Hence for the companied implementing blockchain focus should be on sharing success stories of different companies with the advantages of implementing of blockchain as opposed to using traditional methods.

Lack of regulations and counter party fraud was found to have a significant impact on blockchain adoption. This means that the practitioners are somewhat skeptical about using the technology due to lack of regulations which is the reason why Blockchain has not been implemented yet in large scale. Further, it also reflects that the practitioners perceive a level of distrust towards the blockchain technology, as they do perceive potentially harmful consequences of blockchain technology. Further, in order to accommodate the implementation of blockchain technology, respective organizations will face costs originating from: hiring subject matter experts, data scientists, educational programs etc. which are key in aiding the implementation of blockchain in an organization. Hence the factors studied in this paper has a significant effect and must be accounted for when determining the cost to benefit ratio of implementing blockchain technology.

References

A. Dorri, M. Steger, S. S. Kanhere and R. Jurdak, "BlockChain: A Distributed Solution to Automotive Security and Privacy," in IEEE Communications Magazine, vol. 55, no. 12, pp. 119-125, Dec. 2017.

Atzori, M. (2017). Blockchain Technology and Decentralized Governance: is the State Still Necessary? Journal of Governance and Regulation, 6(1), 45-62.

Bal, M. (2017). Securing property rights in India through distributed ledger technology. New Delhi: Observer Research Foundation.

Bass, J. (2019). The Truth About Blockchain and Its Application to Health Care. Hfm (Healthcare Financial Management), 1–9

Bhattacharyya, S. S., & Nair, S. (2019). Explicating the future of work: perspectives from India. Journal of Management Development, 38(3), 175-194.

Bhattacharyya, S. S., & Nair, S. (2019). Explicating the future of work: perspectives from India. Journal of Management Development, 38(3), 175-194.

Deloitte (2017). Blockchain technology in India Opportunities and challenges

Folkinshteyn, D., & Lennon, M. (2016). Braving Bitcoin: A technology acceptance model (TAM) analysis. Journal of Information Technology Case and Application Research, 18(4), 220-249.

Gainsbury, S. M., & Blaszczynski, A. (2017). How blockchain and cryptocurrency technology could revolutionize online gambling. Gaming Law Review, 21(7), 482-492.

Gartner (2019). Customer-Driven Alliances Will Unlock Blockchain Potential

Gupta, A., & Gupta, S. (2018). Blockchain Technology: Application in Indian Banking Sector. *Delhi Business Review*, *19*(2), 75-84.

Grover, P., Kar, A. and Janssen, M. (2019), "Diffusion of blockchain technology", Journal of Enterprise Information Management, Vol. 32 No. 5, pp. 735-757.

Guych Nuryyev , Yu-Ping Wang , Jennet Achyldurdyyeva , Bih-Shiaw Jaw , Yi-Shien Yeh , Hsien-Tang Lin and Kamble, S., Gunasekaran, A., & Arha, H. (2019). Understanding the Blockchain technology adoption in supply chains-Indian context. International Journal of Production Research, 57(7), 2009-2033.

Kamilaris, A., Fonts, A., & Prenafeta-Boldú, F. X. (2019). The rise of blockchain technology in agriculture and food supply chains. Trends in Food Science & Technology, 91, 640-652.

Knauer, F., & Mann, A. (2019). What is in it for me? Identifying drivers of blockchain acceptance among German consumers. The Journal of the British Blockchain Association, 10484.

Kamble, Sachin & Gunasekaran, Angappa & Arha, Himanshu. (2018). Understanding the Blockchain technology adoption in supply chains-Indian context. International Journal of Production Research. 1-25. 10.1080/00207543.2018.1518610.

Li-Fan Wu, 2020. Blockchain Technology Adoption Behavior and Sustainability of the Business in Tourism and Hospitality SMEs

Lee, C. C., Kriscenski, J. C., & Lim, H. S. (2019). AN EMPIRICAL STUDY OF BEHAVIORAL INTENTION TO USE BLOCKCHAIN TECHNOLOGY. *Journal of International Business Disciplines*, *14*(1).

Nasscom (2019). NASSCOM Avasant India Blockchain Report 2019

Sadouskaya, K. (2017). Adoption of Blockchain Technology in Supply Chain and Logistics.
Schuetz, Sebastian & Venkatesh, Viswanath. (2019). Blockchain, adoption, and financial inclusion in India: Research opportunities. International Journal of Information Management. 10.1016/j.ijinfomgt.2019.04.009.

Sheel, A., & Nath, V. (2019). Effect of blockchain technology adoption on supply chain adaptability, agility, alignment and performance. Management Research Review.

Siba, T. K., & Prakash, A. (2016). Block-Chain: An Evolving Technology. Global Journal of Enterprise Information System, 8(4), 29–35.

Suh, B., & Han, I. (2003). The impact of customer trust and perception of security control on the acceptance of electronic commerce. International Journal of electronic commerce, 7(3), 135-161.

Shah, T., & Jani, S. (2018). Applications of blockchain technology in banking & finance. *Parul CUniversity, Vadodara, India*.

Varma, J. R. (2019). Blockchain in Finance. Vikalpa: The Journal for Decision Makers, 44(1), 1–11.

DECISION SCIENCES INSTITUTE

Blockchain Applications: SME Interviews and Financial & Banking Use Case

Van Tran
University of Northern Iowa
Email: vtran@uni.edu ; haivantran.29@gmail.com

Matt Barton
Epic
Email: bartomab@uni.edu ; mrb708058@gmail.com

Dan Bumblauskas, Ph.D.
University of Northern Iowa
Email: daniel.bumblauskas@uni.edu

Arti Mann, Ph.D.
University of Northern Iowa
Email: arti.mann@uni.edu

ABSTRACT

Blockchain is an emerging technology that is believed to be “the next internet.” In this paper, we try to answer the questions- What is blockchain? How does blockchain work? and what are its uses across different industries? We also explore blockchain from a theoretical perspective and discusses its various applications. We analyze the drawbacks of this technology and also explore the uses of this technology in banking security and risk management.

KEYWORDS: Blockchain, Cybersecurity, Bitcoin, Digital Identity Management, Supply Chain

INTRODUCTION

Throughout human history, few technological advancements have so profoundly impacted the future that they are considered general-purpose technologies. Among the ranks of these advancements include the steam engine, electricity, and robotic automation (Muro & Andes, 2015). Blockchain, a driving force behind cryptocurrencies and simultaneous transactions, is predicted by many experts to emerge as the new general-purpose technology (Bullock, personal communication, 2021; [removed for blind peer review]). The adoption of blockchain was further pushed by the COVID-19 pandemic, as it revealed weaknesses in current supply chains and data management systems. Organizations, government agencies, and companies partnered to build blockchain-based solutions to effectively map and analyze the medical supply chain (van Hoek & Lacity, 2020). Blockchain was also used to build a faster and more secure data-sharing system amongst individuals, hospitals, governments, and other organizations (van Hoek & Lacity, 2020). The COVID crisis created the perfect storm for hackers around the world as they took advantage of the chaos and the expanding technology footprints and attack surface. The finance sector became one of the main targets as finance-related attacks accounted for 52% of all attacks seen across the VMware Carbon Black dataset, which was ‘an unprecedented anomaly in [their] data tracking’ (Upatham & Treinen, 2020). To combat the increasingly sophisticated frauds and cyberattacks, financial companies started looking into integrating blockchain into their security systems. Blockchain's ability to foster data sharing without affecting data privacy and security can help improve data-driven decision-making processes while maintaining democratic values.

This brings us to the questions: What is blockchain?, How does blockchain work?, and What is blockchain used for? Through a careful analysis of publications, expert opinions, and informed research, this article defines blockchain, explore its theories, and examine its uses. We will also explore the concepts of the so-called double-spending problem, as well as the immutability and distributed ledger properties that solve this problem and so clearly define blockchain. We further provide a discussion on the uses of blockchain and focus on specific fields where blockchain may prove especially useful, such as in trade between new partners, agriculture and food, finance, healthcare, and security. Drawbacks of blockchain are also explored, including environmental impact and a lack of regulation so notable that it is even referred to as the "wild west" by experts (Bullock, personal communication, 2021; [removed for blind peer review]), will also be investigated. Finally, an anticipated future of blockchain applications will be provided.

Through this paper, a thorough analysis of blockchain technology has been researched. This technology is often regarded as the "next internet" in the business world: while the uses are at times not very apparent, businesses believe the impact of blockchain on the world will be revolutionary. As such, this paper will explain how blockchain works from a theoretical perspective, exploring the concepts and foundations that enable its functionalities as well as the current uses and applications of blockchain with a focus on banking security and risk management. This paper also serves to explain the drawbacks of this technology, as well as possible threats to its future. Finally, using this information, this work will answer the question: what applications will blockchain likely support in the future?

LITERATURE REVIEW

In its simplest form, a blockchain is a continuous ledger of business transactions that can never be deleted; the ledger can only be altered by adding additional transactions (Mougayar, 2016). It removes 'middlemen' entirely or requires them to add new information in blockchain processes because "trust" can be found by simply following the blockchain trails (Mougayar, 2016). In other words, middlemen, such as brokers who find trustworthy trade partners, will no longer be needed to determine how trustworthy another party is because their history is documented in the blockchain. Furthermore, while traditional databases are administered and only create the "illusion" of reliability, blockchain technology is cataloged such that the data is "irrefutable" (Mougayar, 2016). On a similar note, blockchains can be used to confirm a transaction, and this confirmation means that it is less likely that a transaction will be canceled (Ricci et al., 2019). Blocks are reconfirmed each time a block is added to the blockchain, meaning that all prior transactions become less likely to be reverted whenever a new block is added (Ricci et al., 2019). Blockchain can create an efficient and trustless environment through smart contracts, which are unbiased, digitized agreements that automatically execute if and only when all requirements of a transaction are met. Thus, one can safely conclude that, with blockchain, there are benefits to businesses that conduct business in a trustworthy manner, lest they have a long, permanent history of reverted transactions or, even worse, an undeniable history of disreputable behavior documented in the blockchain.

There are two types of blockchains – permissioned (private) and permissionless (public). A public blockchain is open to everyone, and an example of a public blockchain is the platform that runs Bitcoin. Public blockchain platforms employ a crypto-economic model that rewards users' tokens as a way to incentivize them to contribute to the networks (Kadiyala, 2018). The more decentralized and distributed the network is, or the longer the transaction history, the harder it is to hack into. On the other hand, a permissioned blockchain is only accessible to a certain number of nodes who have permission to enter. It is usually built for an organization or a consortium to exchange information. A private blockchain is not completely decentralized as the identities of all nodes are managed. The consensus mechanisms of these blockchains are relatively inexpensive when compared to the public blockchain. They are also much more scalable and efficient than permissionless platforms (Kadiyala, 2018). As they do not follow the same crypto-economic model, many permissioned blockchains do not have their currency.

The value of blockchain, as well as its applications, are founded on five basic principles: “truth and trust,” “transparency,” “security,” “quality and certainty,” and “efficiency” (Welfare, 2019, p. 10-11). Welfare explained that one major use of blockchain is the utilization of cryptocurrency, which are essentially virtual currencies that demonstrate the same characteristics as other blockchain technologies (permanent transaction records, distribution of these records, and all other characteristics). Welfare goes on to explain applications in “identity management”, “loyalty” programs, “warranty management and refund management”, along with numerous supply chain applications such as the handling of inventory, logistics and transportation, and even counterfeit prevention. Bumblauskas et al. (2020) explores the applications utilized by Bytable Inc. for the distribution of eggs. Specifically, this study uses blockchain to determine the routes eggs take as they travel “from farm to consumer.” Blockchain enables a variety of benefits to the egg industry. First, it allows for compliance with FDA regulations. Second, when utilized with other technologies, relevant information can be easily maintained. In the egg use case, temperature sensors are used to add temperature information to the blocks in the blockchain so that Bytable, Inc. knows the eggs were stored at a safe temperature. In the event of recalls, Bytable Inc. will know which sources required a recall and will know exactly which egg cartons came from which sources because it is all stored in the blockchain. Finally, “food fraud and ethics” are accounted for because Bytable, Inc. can ensure their suppliers uphold the ethical standards their consumers are willing to pay extra for (e.g., fair treatment of the chickens that lay the eggs) (Bumblauskas et al., 2020). Ultimately, Bumblauskas et al. (2020) determined blockchain is a viable means of food tracing.

Regarding future research, Ante (2020) identifies further avenues for research to be conducted in the field of blockchain. Ante (2020) considers the original theories for blockchain presented by Satoshi Nakamoto in 2008 and seeks additional room for research in his article. In addition to applications of blockchain, the area of research this study will explore, Ante (2020) finds a need to conduct further research into the anonymity of transactions. Blockchain allows for anonymous transactions if the information linking the transaction to one specific person is withheld, as explained by Nakamoto (2008) in the original white paper explaining the concept of Bitcoin.

METHODOLOGIES

The first aspect of the methodology for this study involves the analysis of scholarly articles and other publications on the subject matter. Through analysis and synthesis of these publications, strong background details and a preliminary summary of the applications of blockchain, both current and future, will be produced.

The next aspect of the methodology behind this study is a small set of interviews with experts on the subject matter. The first person interviewed is Bert Bullock, a long-time expert in computer technology with a rich history of introducing and utilizing cutting-edge technology in various businesses including Alcare Computers, EDS, AtlasVac, and other businesses in the technology sector. Bert regularly studies the latest innovations in regards to information systems, and blockchain is one such technology he studies. The second person interviewed [has been removed for blind peer review].

Finally, concepts and current applications of blockchain are synthesized into a projection of the future of blockchain and the businesses that will be most impacted by these innovations. This synthesis incorporates ideas proposed by the aforementioned experts, as well as utilizes a working knowledge established through a review of literature and publications on blockchain technology. These projections are not merely wishful thinking and personal predictions, but rather a well-defined and highly plausible image of the future of blockchain.

THEORIES BEHIND BLOCKCHAIN

The greatest aspects of the blockchain theory lie in the original Bitcoin white paper, written by the enigmatic Satoshi Nakamoto, which presents the first recommended solution to an issue known as the double-spending problem. As Entertainment Newsweekly (2020) describes, “in blockchain networks, due to reproducibility of data, a digital asset could be reused. For example, multiple parallel blockchain transactions can be performed on the same (sic) digital asset (referred to as double-spending).” This, of course, presents a very clear counterfeiting problem, as digital assets such as digital currencies would only hold value if the assets cannot be replicated freely. Nakamoto (2008) produces “a solution to the double-spending problem using a peer-to-peer network” in Bitcoin. Bitcoin, being confirmed in this peer-to-peer system, prevents double-spending; it uses hashing to prove the order in which transactions occur, and it utilizes cryptography to prevent attackers from modifying this chronological data. By proving the chronological order of transactions, the blockchain would not accept multiple transactions to occur on one asset.

This system, of course, depends on a peer-to-peer network, as Nakamoto (2008) describes. Because many transactions take place in blockchain, a large number of computers are needed to constantly confirm the order of transactions. The computers not only prove the transactions but encrypt the data in the blocks so that an attacker cannot later alter the data to enable double-spending, according to Nakamoto (2008). The encryptions used in Bitcoin, as Bullock (personal communication, 2021) describes, are incredibly difficult to break, taking unfathomably long amounts of time for a classical computer to solve. For this reason, blockchain takes on its immutability: if the blocks could be modified, the double-spending problem would remain unsolved, as attackers could change the data and double-spend the digital assets.

Chronology, however, is not enough to completely prevent the double-spending problem. Nakamoto (2008) recognizes that there must also be a system to identify which assets are legitimate and which are simply replicas. Therefore, a repository documenting which assets are real and which assets are fraudulent proves necessary, but Nakamoto (2008) did not want to use a central “mint”, as he called it, that users would simply have to trust. Thus, the distributed ledger emerges: if all computers engaged in the network could agree on which assets and transactions are legitimate, then any outside fraudulent assets used by attackers attempting to double-spend could be recognized as fakes (Nakamoto, 2008). The distributed ledger quite literally distributes the ledger across every computer in the network, meaning all computers engaged in the blockchain processes must store the entire blockchain. To be clear, as Nakamoto (2008) described blockchain, there is no location where any data should be stored or accessed beyond blockchain participants' computers. This adds an additional layer of security: no attacker can simply access a central data repository and change information, as there is no repository to attack. Therefore, even if one node was altered, the chain would be protected. One obvious concern, of course, relates to the storage of all this data, but Nakamoto (2008) dispels these concerns. The storage space required to hold every single block generated in a given year, assuming new blocks are generated every ten minutes, will on average require "4.2 MB (megabytes) per year (Nakamoto, 2008)." For perspective, some of the smallest hard drives one could reasonably expect to find in a modern computer hold over 30,000 times as much data.

The entire theory enabling blockchain solutions that do not double-spend are summarized in a set of steps, also described by Nakamoto (2008) in his white paper. To summarize, somebody attempts to complete a transaction, and all computers in the network are notified. Then, one "node" of the network "collects new transactions into a block (Nakamoto, 2008)." Cryptography is applied to secure the data in the transaction, then the original node informs all the other nodes within the network. All other computers in the network then verify with each other that "all transactions in (the block) are valid and not already spent (Nakamoto, 2008)." Finally, the transaction is solidified by locking it in place when a new block is added to the chain. Through these steps, Nakamoto's theoretical blockchain that solves the double-spending problem can be expressed in all practical applications of blockchain.

APPLICATIONS OF BLOCKCHAIN

On its most foundational level, blockchain creates trustworthiness in transactions. As Bert Bullock (personal communication, 2021) explains, blockchain allows for "simultaneous" transactions in which ownership titles and currency are exchanged at precisely the same moment in time. As a result, the transaction is added to the blockchain, all items in the transaction immediately change hands, and there is no need to trust the other parties in the transaction. As Bullock (personal communication, 2021) explains, "blockchain builds trust where trust is in dispute."

An area of blockchain application that has been gaining a lot of interest in recent years is fintech. Despite blockchain being a fairly new technology, the benefits that it brings are being widely recognized by business leaders in the finance sector. According to a survey conducted by Cognizant - a technology company that specializes in business consulting, information technology and outsourcing services - firms in the financial sector believe improved data management, improved risk management, heightened security, reduced fraud, and improved auditing are among the top benefits of blockchain (Varghese et al., 2017, p. 8). A study by Taylor et al. (2020) shows that the majority of blockchain research on cybersecurity focuses on the Internet of things (IoT) (45%) and Data storage and Sharing (16%) (p. 150). Leaders in the industry are starting to look into blockchain to create a secured environment that can protect them against cyber-attacks and frauds.

Firstly, the implementation of blockchain can help banks control risks much more effectively. All transactions being encrypted, tamperproof, time-stamped, and tracked in real-time discourage fraudulent activities. As a result, communications among peers are protected, and the integrity of data is ensured. In addition, audits can be conducted more effectively. Blockchain technology and its transparency element also allow banks to verify the identity of their customers, which saves banks time and money on their Know Your Customer (KYC) process.

Secondly, blockchain can enhance data privacy and security as all data is encrypted and stored across all nodes in the network instead of just one centralized server, which eliminates a single point of failure. An article by Deloitte discusses the applications of blockchain on the Internet of Things (IoT) security and Distributed Denial of Services (DDoS) attack prevention. A DDoS attack overloads a targeted server by recruiting multiple computers (botnets) connected to the Internet to send traffic simultaneously and repeatedly. Hackers either remotely access devices using easily guessable login credentials to install malware, or they launch DDoS attacks with a Command and Control server, which is a master server that gives instructions for the bots to read and act on (Sallaba et al., 2017, p. 2). Instead of the default login credential, blockchain requires devices to use a public key and private key cryptography that would only be known to the user, making the system more difficult to hack (Sallaba et al., 2017, p. 2). As a decentralized, peer-to-peer network, the attacker's Command and Control server will not be able to gain access to control other nodes to launch a DDoS attack. Sallaba et al. also notes that blockchain can eliminate the use of Domain Name System (DNS) server, which is a centralized server that maps IP addresses to domain names (2017, p. 2-3). As the name and address pair will be stored on blockchain and copied across all nodes, there is no longer a need for a DNS server and therefore, no one single point of failure. Lastly, the distributed and shared nature of the blockchain could facilitate the recovery of both data and processes in the case of an attack (assuming that not all the nodes are corrupted simultaneously). This could reduce the need for costly recovery plans (Dzhaparov, 2020, p. 47).

As smart contracts are unbiased, digitized agreements that only execute when all requirements are met, less trust is needed between partners. This, along with other benefits of blockchain discussed above, provides a safe environment that encourages the sharing of information between partners, with or without trust being established in advance.

As mentioned above, the use of blockchain in digital identity management is one of the areas that are gaining interest. Since the COVID-19 pandemic moved many different financial services online, banks need to rethink their identity verification and credential sharing processes as physical contacts need to be kept to a minimum. Sensitive data is stored in centralized databases by third parties opens up many vulnerabilities for hackers to exploit. Furthermore, customers having to manage multiple accounts and passwords over many different websites also put their data at risk. They can forget their log-in credentials, which in return might force them to provide third parties with even more personal data to regain their passwords. Similar, easier-to-remember passwords can also be used over multiple websites as a result, leaving many of their online accounts vulnerable. Lastly, the KYC policies have always been considered overly costly and complex.

Blockchain introduces a new identity management system with the concept of self-sovereign identity, which allows an individual (a person or an organization) to be in full control and ownership of their own data. Figure 1 is a blockchain's identity management framework proposed by Aydar, Ayvaz, and Cetin (2020, p. 7-17).

Figure 1 *Overall workflow of the proposed identity system*

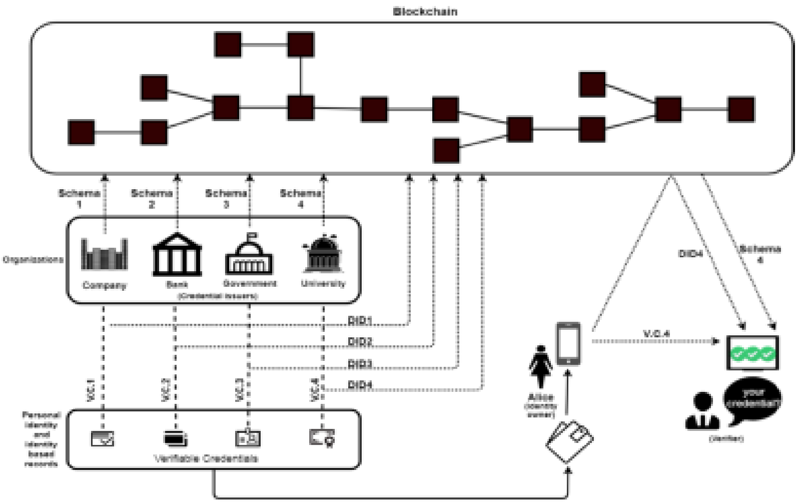


Figure 1 Note. From “Towards a Blockchain based digital identity verification, record attestation and record sharing system,” by M. Aydar, S. Ayvaz and S. Cemil, 2020, p. 8 (https://arxiv.org/abs/1906.09791). arXiv.org perpetual, non-exclusive license.

In the system above, a digital identity is assigned to each individual through blockchain. The blockchain itself does not store any actual private personal data, but rather the proof of verification (Aydar et al., 2020, p. 8). For example, if someone’s driving license is being used to verify their identity, the actual license is not stored on the blockchain. What will be on the blockchain is just the fact that their document has been verified by, in this case, the local government and therefore is good to use (Aydar et al., 2020, p. 8). Users can store all of their verifiable credentials, which are machine readable, cryptographically secure digital credentials, in an identity wallet. Once again, identity owners fully control their data and can divide whom to share their data.

This identity management system allows banks to streamline their customer on-boarding processes. Banks no longer have to spend as much time, money and labor on verifying customers’ information. Transparency as well as information accuracy and reliability are also significantly increased. Due to the secure nature of blockchain, the risk of ID theft and fraud is also reduced.

For customers, blockchain’s identity management system provides a much more convenient user experience as all verifiable credentials are stored in one place, users no longer need to memorize multiple log-in credentials or reach out to third parties whenever they need to verify their identities. Customers are once again in full control of their own data, which provides full transparency between customers and financial institutions.

However, this framework is not without flaws. Since all identities are essentially created and only exist in the digital world, banks still need to figure out who is responsible to provide the trust mapping between real life physical identity and the digital identity (“Practical thoughts,” n.d., p. 9). It will be crucial to ensure that a vulnerable identity provider doesn’t open up opportunities for an identity takeover on the network. Furthermore, with blockchain still being a very new technology, businesses may face resistance from both partners and their own employees. In order to retain their customers, verifying organizations may refuse to share data or collaborate

(“Practical thoughts,” n.d., p. 10). Lastly, there has not been many clear regulations on the use of blockchain in the financial sector. As a result, all entities involved have to accept risk and uncertainty by agreeing to participate in an identity network.

Figure 2 shows an example use case of a loan application (Aydar et al., 2020, p. 11). In this use case, the customer Alice, who is a current customer of Bank A, wants to apply for a loan with Bank B. Bank A and Bank B are trusted partners. Alice also owns land, and that information is kept by the local government. Because Bank B has never worked with Alice before, they are required by KYC regulations to know her before accepting her application. Using the proposed framework, Allice authenticates herself with Bank B, who then reaches out to related organizations to collect the information needed for the loan application. Alice will need to give her consent to these organizations before they can share her information in the form of verifiable credentials. The whole process happens within minutes without physical interaction between any parties involved.

Figure 2 *Use Case: Loan Application*

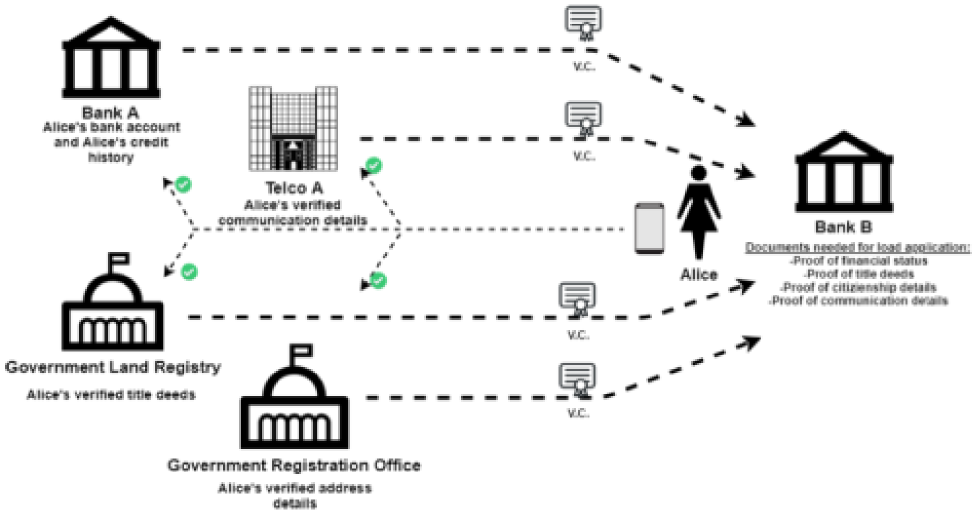


Figure 2 Note. From “Towards a Blockchain based digital identity verification, record attestation and record sharing system,” by M. Aydar, S. Ayvaz and S. Cemil, 2020, p. 11 (https://arxiv.org/abs/1906.09791). arXiv.org perpetual, non-exclusive license.

DRAWBACKS AND LIMITATIONS OF BLOCKCHAIN

Blockchain, despite its many advantages, is not without its risks and downsides. Concerns are especially strong in regards to the tremendous environmental impact in processes that do not add value to the blockchain, which are derived from significant energy usage. Another major concern is the lack of regulation and possibilities for hacking, allowing so-called crypto-heists to occur. Numerous other issues, which have received less research attention, also exist and will be discussed as well. For blockchain to become a more viable solution, these concerns must be addressed.

Blockchain's dependence on peer-to-peer computer networks and cryptography presents a strong disadvantage for the environment. As Bullock (personal communication, 2021) explains, tremendous costs are necessary to add each block to a blockchain. Bullock is not alone, as Juričić et al. (2020) acknowledge that "blockchain is still strongly criticized for its lack of usefulness and resource-heavy consumption." These authors continue to explain that, in this blockchain system, computers must complete large amounts of calculations that add no direct value at all, but merely allow the system to function accurately. Juričić et al. (2020) describe using a graph that bitcoin mining, which is the process that validates blocks, uses approximately 75 trillion watt-hours of energy every year, and is estimated to soon consume nearly as much energy as the country of Ireland in these processes that add no direct value to any tangible goods or services. Juričić et al. (2020) further explain that most of this mining occurs in places that burn coal to produce power, generating a great deal of pollution in the process.

Another serious and concerning aspect of blockchain, in its current state, is its distinct lack of regulation. This lack of regulation serves as a double-edged sword, as Bullock (personal communication, 2021) explains: while there is much room for innovation to occur unhindered, there is also no authority to turn to when things go wrong. Botsman (2017) identifies one particular case where the Decentralized Autonomous Organization (DAO), which was based on a blockchain network known as Ethereum, was hacked and nearly lost over 60 million US dollars in cryptocurrency. Botsman (2017) further explains that rules were written into the Ethereum network's code to reverse the transaction using a "hard fork," which is "technical jargon for essentially rewriting history or changing the rules." Nonetheless, the DAO had no laws or governments to turn to and had to make this decision for themselves. Their decision to use a hard fork, Botsman (2017) documents, cost significant trust in the Ethereum network by performing what many considered a "blockchain sin," as it ran counter to numerous blockchain principles set forth by Nakamoto (2008) in his white paper, particularly in regards to the finality and immutability of transactions.

In addition to these major disadvantages of blockchain, Niranjanamurthy et al. (2019) and Bullock (personal communication, 2021) highlight other, less-researched drawbacks of this technology. Bullock expresses concerns about the amount of time required by a computer to validate blocks, and these concerns are echoed by Niranjanamurthy et al. (2019). This time is primarily derived from the time it takes to complete cryptographic processes, as well as time spent updating nodes in blockchain's peer-to-peer network. In his paper "The Quest of Scalable Blockchain Fabric," Marko Vukolic brings forth the tradeoff between scalability and performance amongst the two types of blockchain - Proof-of-Work (PoW) based and Byzantine fault-tolerant (BFT) based (2015, p. 4). PoW blockchain is the technology behind Bitcoin which requires nodes to "mine," or to add another block to the chain. A block is usually added every ten minutes, and mining takes up a lot of computing power (Vukolic, 2015, p. 2). However, PoW blockchain is a public blockchain and offers good scalability. On the other hand, BFT-based blockchain offers good performance for a small number of replicas or nodes (Vukolic, 2015, p. 3). This brings up the discussion of the Scalability Trilemma, a term coined by the founder of the public blockchain Ethereum, Vitalik Buterin (Viswanathan & Shah, 2018). The trilemma refers to the tradeoffs between security, scalability, and decentralization. Currently, there are no blockchain platforms that can optimize all three of these factors, and it is up to the business to decide which factors are best suited for their needs.

In addition, Niranjanamurthy et al. (2019) also acknowledge concerns in integrating blockchain systems, which inherently require the replacement of existing systems. As a result, implementing a blockchain framework into a new company requires both time and money, and it requires “cultural adoption,” all of which can prove difficult to overcome. It is also worth noting that blockchain technology has only been around for ten years. Smart contracts, though more reliable, are expensive and difficult to program. Several studies selected analyzed proposals or concepts and have little quantitative data or practical applications.

In addition, Niranjanamurthy et al. (2019) identify further hacking risks besides the aforementioned crypto-heists, including “user identity theft,” “injection of malicious code into a distributed ledger,” and “fictitious blockchain applications (that) will appear to steal transaction details/personal information/behavior from nodes/individuals,” among others (pp. 14754-14755). While blockchain’s decentralization has the power to transform governments and businesses, it introduces a different set of weaknesses. For industries that require a high level of confidentiality such as financial services, private data can be monitored and pierced together to become a serious data privacy breach as all data is being shared with all users on a network (Fitzpatrick, 2019).

THE FUTURE OF BLOCKCHAIN

Blockchain, being in its infancy, has only scratched the surface of its great potential thus far. Indeed, as Ante (2020) describes, room exists to further study the practical applications of blockchain. Based on the insight of experts in the field of blockchain, in addition to well-informed personal speculation, a glimpse into the future of blockchain can be found.

As Bullock (personal communication, 2021) explains, blockchain resolves issues where trust is questionable, which illustrates the strong potential for blockchain to facilitate contract negotiations. According to Bullock (personal communication, 2021), these applications are less developed than other existing applications, such as cryptocurrency. He explains that, because blockchain can allow simultaneous transactions when certain conditions are met, blockchain offers a contract solution by guaranteeing all parts of the transaction, such as funds and ownership titles, are traded at precisely the same time, removing the ability for one side to con the other during the trade. Furthermore, because of the immutable nature of blocks, blockchain may prove incredibly powerful at resolving contract-based conflicts, as the evidence cannot be tampered with. Simply put, Bullock (personal communication, 2021) envisions blockchain will serve as a “data repository to hold the contractual obligations between the parties.”

Both Bullock (personal communication, 2021) and [removed for blind peer review] (personal communication, 2021) expect blockchain to manifest itself especially prominently in procurement, namely regarding bills of materials. Bullock (personal communication, 2021) describes a system where blockchain facilitates just-in-time delivery utilizing the same simultaneous transaction mechanisms as in contracts. When certain criteria are met, such as inventory dipping below a certain threshold and a sufficient number of parts becoming available from a supplier, blockchain can facilitate automatic transactions to maintain an optimal supply of inventory. [removed for blind peer review] (personal communication, 2021) suggests that a different use for blockchain in bills of materials may also emerge. He suspects that a “blockchain button ([removed for blind peer review])” may be incorporated into applications so that users can see details about their product that are currently unavailable, such as where each part in the bill of materials was sourced and the steps the product took to ultimately arrive in the hands of purchasers. Such an application would especially benefit socially conscious consumers and companies who may wish to know these smaller details, such as how environmentally friendly a product labeled as “organic” actually is or if products were manufactured in facilities with a positive track record for treatment of employees.

Another potential phase blockchain may face is centered around the idea of standardization. Because blockchain is still within early phases of development, there are multiple different blockchain systems in use at the same time, some of which are entirely incompatible with each other ([removed for blind peer review]). Considering the concerns about the energy consumption of blockchain, [removed for blind peer review] (personal communication, 2021) recognizes that “there are different ways to create (blockchain) ledgers that are less energy-intensive.” He believes that standardization of blockchain is a likely step in the near future to address some of the less-than-desirable outcomes created by the current systems.

CONCLUSION

Blockchain, although early in its life cycle, already demonstrates a powerful capacity to forever change the world of business. The COVID-19 pandemic forced companies in all industries to embrace major technological disruption and further accelerated the development and adoption of blockchain. Specifically, the banking industry - one of the main targets for hackers during the global crisis - gained more and more interest in using blockchain technology to develop a more proactive and advanced security system. Through careful analysis of articles, literature, and other written works, as well as interviews with experts, many foundational concepts of blockchain have been explained. Blockchain’s most essential aspects include a distributed ledger, transparency, and immutability, and the theories behind blockchain solutions that prevent double-spending help guarantee these qualities remain in effect. Blockchain demonstrates applications today as a trust-building tool that can provide a decentralized, trustless and transparent platform that can significantly improve security, data management and risk management in the finance sector as well as other industries. This paper introduces a blockchain-based digital identity framework that promises to foster collaboration and data sharing between businesses and their partners without the need for physical interactions between parties involved. Nonetheless, blockchain is not without its disadvantages, as it consumes tremendous amounts of power, proves to contain some vulnerabilities to malicious actors, requires a great deal of validation time for blocks, and presents organizational risks in implementation and adoption. Despite these disadvantages, blockchain has a bright future, with the most likely future applications appearing in contract negotiations and bills of materials. With these applications, and because blockchain is largely unregulated, future standardizations are also expected to come into existence. While blockchain is still early in its development, make no mistake: the next general-purpose technology has arrived.

REFERENCES

Ante, L. (2020). A place next to Satoshi: foundations of blockchain and cryptocurrency research in business and economics. *Scientometrics*, *124*(2), 1305–1333. https://doi.org/10.1007/s11192-020-03492-8.

Aydar, M., Ayvaz, S., & Cemil Cetin, S. (2020, June 23). Towards a blockchain based digital identity verification, record attestation and record sharing system. *arXiv e-prints, arXiv-1906.*

Botsman, R. (2017). The Great Crypto Heist. The Australian Financial Review, pp. The Australian financial review, 2017-10-13.

Bullock, B. (2021, March 9). Personal communication [Personal interview].

Bumblauskas, D., Mann, A., Dugan, B., & Rittmer, J. (2020). A blockchain use case in food distribution: Do you know where your food has been? *International Journal of Information Management*, *52*.

Dzhaparov, P. (2020). Application of blockchain and artificial intelligence in bank risk management. Икономика и управление, 17(1), 43-57.

Entertainment Newsweekly (2020). "Method And Device For Avoiding Double-Spending Problem In Read-Write Set-Model-Based Blockchain Technology" in Patent Application Approval Process (USPTO 20200349568). (2020, November 27). Entertainment Newsweekly, 626. https://link.gale.com/apps/doc/A642477304/ITOF?u=uni_rodit&sid=ITOF&xid=0fef0cee

Fitzpatrick, L. (2019, February 2). *A hacker's take on blockchain security.* Forbes. https://www.forbes.com/sites/lukefitzpatrick/2019/02/04/a-hackers-take-on-blockchain-security/.

Juričić, Vedran, Radošević, Matea, & Fuzul, Ena. (2020). Optimizing the Resource Consumption of Blockchain Technology in Business Systems. *Business Systems Research, 11*(3), 78-92.

Kadiyala, A. (2018, February 17). *Nuances between permissionless and permissioned blockchains*. Medium. https://medium.com/@akadiyala/nuances-between-permissionless-and-permissioned-blockchains-f5b566f5d483.

Mougayar, W. (2016). *The business blockchain: promise, practice, and application of the next Internet technology*. John Wiley & Sons, Inc.

Muro, M., & Andes, S. (2015, June 16). Robots Seem to Be Improving Productivity, Not Costing Jobs. *Harvard Business Review*.

Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. bitcoin.pdf. https://bitcoin.org/bitcoin.pdf.

Niranjanamurthy, M, Nithya, B N, & Jagannatha, S. (2019). Analysis of Blockchain technology: Pros, cons and SWOT. Cluster Computing, 22(S6), 14743-14757.

Practical thoughts on blockchain and identity. Okta. (n.d.). https://www.okta.com/resources/whitepaper/practical-thoughts-on-blockchain-and-identity/.

Ricci, S., Ferreira, E., Menasche, D. S., Ziviani, A., Souza, J. E., & Vieira, A. B. (2019). Learning Blockchain Delays. *ACM SIGMETRICS Performance Evaluation Review*, *46*(3), 122–125. https://doi.org/10.1145/3308897.3308952.

Sallaba, M., Kumar, S., & Paulsen, J. H. (2018, December 03). *Prevention of DDoS attacks with blockchain technology.* Deloitte. https://www2.deloitte.com/de/de/pages/technology-media-and-telecommunications/articles/cyber-security-prevention-of-ddos-attacks-with-blockchain-technology.html.

Taylor, P. J., Dargahi, T., Dehghantanha, A., Parizi, R. M., & Choo, K. K. R. (2020). A systematic literature review of blockchain cyber security. *Digital Communications and Networks, 6*(2), 147-156.

Upatham, P., & Treinen, J. (2020, April 15). *Amid COVID-19, global orgs see a 148% spike in ransomware attacks; finance industry heavily targeted*. Carbon Black. https://www.carbonblack.com/blog/amid-covid-19-global-orgs-see-a-148-spike-in-ransomware-attacks-finance-industry-heavily-targeted/.

van Hoek , R., & Lacity, M. (2020, April 27). *How the pandemic is pushing blockchain forward*. Harvard Business Review. https://hbr.org/2020/04/how-the-pandemic-is-pushing-blockchain-forward.

Varghese, L., Tomer, M., & McCraw, F. (2017, June). *Financial services: Building blockchain one block at a time*. Cognizant. https://www.cognizant.com/whitepapers/financial-services-building-blockchain-one-block-at-a-time-codex2742.pdf.

Viswanathan, S., & Shah, A. (2018, October 19). *The scalability trilemma in blockchain*. Medium. https://medium.com/@aakash_13214/the-scalability-trilemma-in-blockchain-75fb57f646df.

Vukolić, M. (2015, October). The quest for scalable blockchain fabric: Proof-of-work vs. BFT replication. In *International workshop on open problems in network security* (pp. 112-125). Springer, Cham.

Welfare, A. (2019). *Commercializing blockchain: strategic applications in the real world*. Wiley.

DECISION SCIENCES INSTITUTE

Individual Differences in the Performance of Stock-Flow Problem: Cognitive Reflection

Mohammed AlKhars
King Fahd University of Petroleum & Minerals
KFUPM Business School
Dhahran 31261, Saudi Arabia
Tel.: +966-13-860-7744
Email: malkhars@kfupm.edu.sa

ABSTRACT

This study aims to investigate whether cognitive reflection can predict the performance of the stock-flow relationship represented by the department store problem. The cognitive reflection test (CRT) was used to measure cognitive reflection and differentiate between people with analytical versus intuitive thinking. A total of 115 graduate and undergraduate students participated in this study. Only 25% and 12% of these educated people, respectively, were able to answer two questions related to the stock-flow relationship. Logistic regression revealed that students scoring high in CRT tended to understand the stock-flow relationship better than those scoring low in CRT.

KEYWORDS: Cognitive reflection test, Department store problem, Accumulation phenomenon, Logistic regression

INTRODUCTION

Dynamic stock-flow (SF) systems comprise three main blocks: stocks that accumulate over time, inflows that increase accumulation, and outflows that decrease accumulation (Cronin & Gonzalez, 2007; Cronin, Gonzalez & Sterman, 2009; Weinhardt et al., 2015; Hendijani, Ghafourian & Attari, 2021). Applications of the accumulation phenomenon are abundant in everyday and business situations. For example, the amount of money in a bank account is affected by deposits and withdrawals. In an operational management context, the inventory of a product in a warehouse increases by purchasing extra units and decreases by selling units. Furthermore, the number of maintenance jobs required increases with the increase in defective machines and decreases as they are fixed. Previous research shows that understanding this concept is not easy for most people (Sterman, 2002; Cronin & Gonzalez, 2007; Cronin, Gonzalez & Sterman, 2009). The department store problem is a classic example that has been used extensively to study different aspects of the SF problem.

Previous research addressed the behavioral aspect of the department store problem and observed the existence of individual differences in performance, which could be predicted by some personality characteristics. Weinhardt et al. (2015) investigated how cognitive reflection and global/local reasoning relate to understanding the SF problem. They concluded that people with a high level of reflective compared to intuitive thinking tend to perform better in SF problems. However, they did not find a significant relationship between global/local reasoning and the SF problem. Hendijani, Ghafourian & Attari (2021) investigated how the rational-experiential thinking style affects SF performance by considering cognitive reflection as a mediator. They found that rational thinking had a positive effect on SF performance while experiential thinking had a negative effect. Furthermore, CRT had a partial mediating effect on rational thinking but did not affect experiential thinking.

Research on the effect of individual differences and personality characteristics on the SF problem is in its initial stages. Since it is important to recruit employees with a high level of understanding of the accumulation phenomenon, more research is required. Weinhardt et al. (2015) stated that people with high levels of analytical thinking tend to perform better in inventory and supply-chain tasks than those with high levels of intuitive thinking. This personality characteristic could be used in the recruitment and promotion processes. However, some replicative research is needed before incorporating such characteristics into the recruiting and promotion system. This study will respond to this call by replicating the study of the effect of cognitive reflection as a predictor of high performance of the accumulation phenomenon.

LITERATURE REVIEW

This section briefly reviews the literature on the SF concept represented by the department store problem and cognitive reflection as a personality characteristic used to predict SF performance.

SF Performance

Real-life dynamic systems can be complex as they incorporate stocks, inflows, outflows, feedback loops, delays, and endogenous flows. Managing such systems is challenging. However, the literature suggests that, even for the simplest form of dynamic systems consisting of stocks, inflows, and outflows, people do not fully grasp the underlying mechanisms (Sterman, 2002; Cronin & Gonzalez, 2007; Cronin, Gonzalez & Sterman, 2009; Weinhardt et al., 2015; Hendijani, Ghafourian & Attari, 2021). Since it was observed repeatedly in several studies, it was called an SF failure.

One explanation for SF failure is that people use a correlation or pattern-matching heuristic when solving SF problems (Cronin, Gonzalez & Sterman, 2009). They assume that there is a linear relationship between stocks and flows, and, therefore, stocks behave similarly to their flows. This is incorrect because the relationship between stocks and flows is not linear. Instead, the net flow, which is the difference between the inflow and outflow, determines the increase or decrease in the stock level.

Recent studies have shown that some factors can improve the performance of SF problems. For example, Baghaei Lakeh & Ghaffarzadegan (2015) found that the performance of the SF task marginally improved when the analytical thinking mode was activated and people answered analytical problems correctly. Moreover, analytical thinking measured by a cognitive reflection test (CRT) is positively related to SF performance (Weinhardt et al., 2015). Other factors affecting SF performance include spatial ability (Veldhuis & Korzilius, 2017) and mathematical knowledge (Qi & Gonzalez, 2015).

Cognitive Reflection

The dual-process theory proposes that the human mind operates under two systems: system 1 and system 2. System 1 is intuitive and spontaneous and provides rapid decisions with minimal deliberation. On the other hand, system 2 is reflective and deliberate and provides thoughtful answers that override the initial answers to the problem (Stanovich & West, 2000). This theory states that, when reflective thinking dominates intuitive thinking, people usually make more precise decisions. Conversely, people who consistently use their intuitive thinking usually make biased decisions. The use of the correlation or pattern-matching heuristic when solving the SF problem is an example of the use of system 1. Most people tend to use system 1 more often while others tend to use system 2.

A CRT was developed to measure individuals' cognitive reflection (Frederick, 2005). The test consists of three questions with initial intuitive answers that are incorrect. However, when people use their analytical thinking (system 2) to override their intuitive thinking (system 1), the correct answers become obvious. CRT results with their intuitive and correct answers are shown in Table 1.

Frederick (2005) found that people scoring low on the CRT tend to choose small and immediate rewards over large and late ones. CRT has also been used in other operations management (OM) contexts. For example, Moritz et al. (2013) found that people scoring high on CRT tend to perform better in the newsvendor problem. These people tend not to use the simple demand chasing heuristic, which people with intuitive thinking usually use to solve the newsvendor problem. Furthermore, Narayanan & Moritz (2015) found that cognitive reflection contributes to the bullwhip effect in the beer distribution game representing a four-stage serial supply chain. Additionally, Moritz, Siemsen & Kremer (2014) found that decision-makers exhibiting a balance between intuitive judgment and cognitive deliberation tend to have lower forecast errors. Moreover, Weinhardt et al. (2015) found that people with high CRT scores tended to solve the SF problem better than those with low CRT scores. Finally, Pan, Shachat & Wei (2020) investigated the relationship between cognitive reflection and inventory management using the economic order quality model. They found that participants scoring high on a CRT tend to earn more profit and adopt more effective policies. Based on this literature review, the following hypothesis is developed:

Hypothesis 1: People using their analytical thinking (system 2) tend to solve the SF problem more accurately than those using intuitive thinking (system 1).

Table 1: The Cognitive Reflection Test with both the intuitive and correct answers

No	Question	Intuitive answer	Correct answer
1	A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?	0.1	0.05
2	If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?	100	5
3	In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half the lake?	24	47

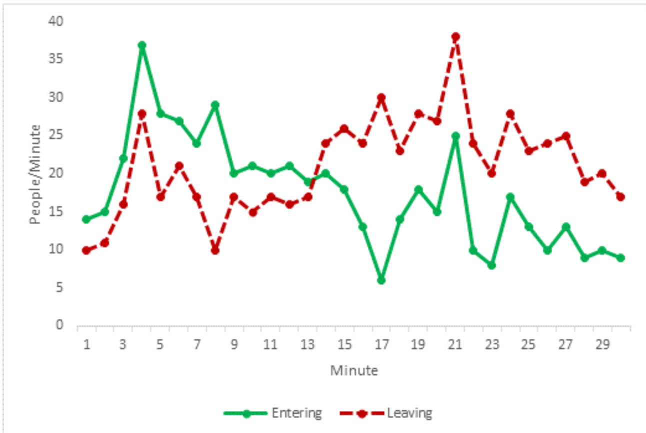
THE DEPARTMENT STORE PROBLEM

The department store problem, shown in Figure 1, has been used to study how people understand the accumulation phenomenon and the relationship between stocks, inflows, and outflows over time. The graph shows the number of people entering and leaving the store each minute over a

30-minute period. Questions 1 and 2 were used to check whether the participant could read the graph correctly. The answer to question 1 is “minute 4” while the answer to question 2 is “minute 22.” Questions 3 and 4 were used to test whether the participant understood the accumulation concept. Question 3 asks, “During which minute were the most people in the store?” while question 4 asks, “During which minute were the fewest people in the store?” The right answer to question 3 is “minute 13.” This is the minute at which the maximum accumulation occurs. Before minute 13, the number of people entering exceeds the number of people leaving. Therefore, the accumulated number of people increases and reaches a maximum of 13 minutes. After minute 13, the number of people leaving the store exceeds the number of people entering the store. Therefore, the accumulated number of people in the store decreases and reaches the minimum at “minute 30,” which is the correct answer to question 4. Cronin, Gonzalez & Sterman (2009) showed that most people could read the graphs well and answered questions 1 (96%) and 2 (95%) correctly. However, only 44% of them were able to answer question 3 correctly, and 31% answered question 4. These results were verified in a subsequent study (Weinhardt et al., 2015) where the percentages of correct answers to the four questions were 86%, 82%, 28%, and 17%, respectively.

Figure 1: Department store problem

The graph below shows the number of people *entering* and *leaving* a department store over a 30-minute period.



Please answer the following questions.

Check the box if the answer cannot be determined from the information provided.

1. During which minute did the most people enter the store?
   Minute __________ ☐ Can’t be determined

2. During which minute did the most people leave the store?
   Minute __________ ☐ Can’t be determined

3. During which minute were the most people in the store?
   Minute __________ ☐ Can’t be determined

4. During which minute were the fewest people in the store?
   Minute __________ ☐ Can’t be determined

METHOD

The survey was distributed to graduate and undergraduate students enrolled in different OM courses. These students were selected because they were expected to have good mathematical and statistical skills. Moreover, some topics in the OM courses dealt with stocks and flows and phenomena such as inventory management. A total of 115 students participated in the survey. The demographic characteristics of the participants are presented in Table 2.

Table 2: Demographic characteristics

	Frequency	%
Academic Level		
Graduate	42	37%
Undergraduate	73	63%
Age		
[18–20]	3	3%
[21–25]	77	67%
[26–35]	32	28%
[36–50]	3	3%
Total	115	

RESULTS AND DISCUSSION

ST Problem

Table 3 shows the number of correct answers to the four questions in the department store problem.

Table 3: Results of the department store problem

	Q1	Q2	Q3	Q4
No	110	107	29	14
%	96%	93%	25%	12%

Most students read the graphs correctly. The percentages of students answering questions 1 and 2 were 96% and 93%, respectively. When it comes to understanding the accumulation phenomenon, only 25% answered question 3 correctly while 12% provided correct answers to question 4. This confirms the findings of previous research (Cronin, Gonzalez & Sterman, 2009; Weinhardt et al., 2015) that people are capable of reading the graph correctly but have difficulty understanding the accumulation phenomenon. Only 12 out of 115 (10.4%) answered the four questions of the SF problem correctly.

CRT

The CRT consisted of three questions. The CRT score for each participant was the summation of the correct answers. Therefore, if a participant answered all questions incorrectly, his/her CRT score was 0. If only one answer was correct, the CRT score was 1. Similarly, if the number of correct answers was two, the score was two. Finally, if all the questions were answered correctly, the CRT score was 3. Table 4 presents the CRT results. The mean CRT score was 1.53.

Table 4: CRT scores				
	0	1	2	3
No	29	28	26	32
%	25	24	23	28

Logistic Regression

Logistic regression was used to test whether cognitive reflection was related to good performance in the accumulation. Two models are presented. The first relates the CRT to the performance on question 3 while the other relates CRT to the performance on question 4. For each model, the dependent variable was whether the answer was correct or incorrect. If the answer was correct, then the value of the dependent variable was 1. If the answer was incorrect, the value of the dependent variable was 0. The independent variable was the CRT score, which was 0, 1, 2, or 3. The results are presented in Table 5.

Table 5: Logistic regression results							
		B	S.E.	Wald	df	Sig.	Exp(B)
Q3	CRT	0.46	0.2	5.37	1	0.02	1.59
	Constant	−1.87	0.43	19.09	1	0	0.15
Q4	CRT	0.58	0.28	4.21	1	0.04	1.78
	Constant	−3.02	0.65	21.49	1	0	0.05

For both models, the CRT was significant in predicting the performance of the department store problem. The B coefficient for question 3 was 0.46 with a p-value of 0.02. Similarly, the B coefficient for question 4 was 0.58 with a p-value of 0.04. The coefficients for the CRT were positive in both models, meaning that people with high CRT scores tended to answer the questions correctly.

CONCLUSION

This study investigated how cognitive reflection as a personality characteristic can be used to predict the performance of the SF phenomenon represented by the department store problem. The results show that highly educated people with extensive training in math and statistics still have difficulty considering the accumulation of people in the department store over time. The research shows that there is a significant relationship between cognitive reflection and understanding of the accumulation phenomenon. People high in cognitive reflection tended to perform better in accumulation problems. This has a practical application in selecting employees for positions that require this type of skill. For inventory-management positions, for example, organizations need employees who understand the accumulation phenomenon. One possible tool for evaluating such candidates is to measure their cognitive reflecting capabilities. It is worth noting that the CRT should not be the only tool used to evaluate candidates for such positions. Research should be devoted to examining whether other personal non-cognitive traits, such as locus of control (Galvin et al., 2018) or impulsiveness (Sharma, Markon & Clark, 2014), can account for the variation in the performance of the accumulation problem. Moreover, the SF phenomenon usually has other dimensions, such as feedback loops, delays, or endogenous flows, contributing to the complexity of the dynamic system. Therefore, future research may investigate how cognitive reflection influences the performance of such complex systems.

REFERENCES

Baghaei Lakeh, A., & Ghaffarzadegan, N. (2015). Does analytical thinking improve understanding of accumulation?. *System dynamics review*, *31*(1-2), 46-65.

Cronin, M. A., & Gonzalez, C. (2007). Understanding the building blocks of dynamic systems. *System Dynamics Review: The Journal of the System Dynamics Society*, *23*(1), 1-17.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational behavior and Human decision Processes*, *108*(1), 116-130.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25-42.

Galvin, B. M., Randel, A. E., Collins, B. J., & Johnson, R. E. (2018). Changing the focus of locus (of control): A targeted review of the locus of control literature and agenda for future research. *Journal of Organizational Behavior*, *39*(7), 820-833.

Hendijani, R., Ghafourian, F., & Attari, I. (2021). The effect of rational-experiential thinking style on stock-flow performance: The mediating role of cognitive reflection. *Current Psychology*, 1-15.

Pan, J., Shachat, J., & Wei, S. (2020). Cognitive reflection and economic order quantity inventory management: An experimental investigation. *Managerial and Decision Economics*, *41*(6), 998-1009.

Moritz, B. B., Hill, A. V., & Donohue, K. L. (2013). Individual differences in the newsvendor problem: Behavior and cognitive reflection. *Journal of Operations Management*, *31*(1-2), 72-85.

Moritz, B., Siemsen, E., & Kremer, M. (2014). Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management*, *23*(7), 1146-1160.

Narayanan, A., & Moritz, B. B. (2015). Decision making and cognition in multi-echelon supply chains: An experimental study. *Production and Operations Management*, *24*(8), 1216-1234.

Qi, L., & Gonzalez, C. (2015). Mathematical knowledge is related to understanding stocks and flows: results from two nations. *System Dynamics Review*, *31*(3), 97-114.

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: a meta-analysis of self-report and behavioral measures. *Psychological bulletin*, *140*(2), 374.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and brain sciences*, *23*(5), 645-665.

Sterman, J. D. (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review*, *18*(4), 501-531.

Veldhuis, G. A., & Korzilius, H. (2017). Seeing with the mind: the relationship between spatial ability and inferring dynamic behaviour from graphs. *Systems Research and Behavioral Science*, *34*(6), 710-727.

Weinhardt, J. M., Hendijani, R., Harman, J. L., Steel, P., & Gonzalez, C. (2015). How analytic reasoning style and global thinking relate to understanding stocks and flows. *Journal of Operations Management*, *39*, 23-30.

DECISION SCIENCES INSTITUTE

Increasing Throughput by Layout Redesign and Resource Allocations for Truck Manufacturing

Goodchild, Duncan
Illinois State University
Email: dgoodch@ilstu.edu

Valiyavetlparambil Sashidharan, Asha
Illinois State University
Email: avaliy1@ilstu.edu

Babatunde, Taiwo
Illinois State University
Email: thbabat@ilstu.edu

Borinara Park
Illinois State University
Email: bpark@ilstu.edu

ABSTRACT

This study introduces a discrete event simulation-based optimization case for increasing manufacturing throughput via layout redesign and employee reallocation. The discrete event method is applied to simulate the operation for re-assigning the employees across different working cells and creating different layouts. The model focuses on utilizing the available resources and layout as factors to increase the throughput in manufacturing. The study concludes the layout changes alone may not create a significant throughput gain but, when combined with the employee reallocation, the layout changes can achieve a significant throughput gain. This type of information would provide a strong case to management to make both changes and in what order.

KEYWORDS: Manufacturing Throughput, Throughput Improvement, Resource Allocation, Process Layout Redesign, Discrete Event Simulation

BACKGROUND

It is estimated the hydro excavating truck global market size will reach $1,450 million USD by 2024, up from US $1,150 million USD in 2019.The industrial truck field of manufacturing contributed an estimated $25.7 billion to the U.S. economy in 2015, with an estimated 225,534 units sold the US alone in 2015 (UC Staff, 2019).

In 2017, over 400,000 excavation accidents were reported in North America, and the pressing need for safer, less invasive equipment could no longer be ignored (DC Velocity Staff, 2017). Hydro excavation trucks offer a more reliable and effective way to excavate as compared to traditional methods of excavation. The growing popularity of hydro excavation started largely in Canada, with it only being a matter of time before the United States embraced the safe and highly efficient technology as well (DC Velocity Staff, 2017).

As demand grew for the safe and labor-efficient vehicles, many companies began to enter the field of hydro excavation truck manufacturing. When it comes to manufacturing, one of the most obvious and most commonly used methods of dealing with an increase in demand is to increase working hours. While it's true that this can help in the short term, it should not be seen as a viable long-term strategy as it does not actually increase the production capacity.

"For manufacturers, throughput in production can mean the difference between meeting quotas and losing customers to the competition." (Steel, 2016). Failing to meet customers' demands, can cause customers to take their business elsewhere or even damage the reputation of the organization. Technological advancements also occur, which lead to increased demand, creating more strain on manufacturers with larger orders. Low productivity usually indicates that resources are not utilized to their maximum potential, which increases company's production costs. In order to increase production capacity and reduce waste, we need to identify problem areas and bottlenecks where the delays occur in the existing process (Gazer, 2019).

LITERATURE REVIEW

An interesting and eye-opening statistic is that most equipment in manufacturing is not utilized anywhere near its true capacity. In fact, the difference between typical manufacturing (with an Overall Equipment Effectiveness score of 60%) and best-in-class manufacturing (roughly an OEE score of 85%) represents up to an astounding 42% increase in capacity (Vorne, n.d.).

In a case study conducted by Gokulraju et al., reducing the lead time and increasing throughput were examined through the use of value stream mapping (Gokulraju et al., 2016). A current state value stream map was created and analyzed for potential areas of improvement, and it was identified that only 6% of the time was value added activities, which means a significant amount in the entire manufacturing time was non-value adding activities.

One of the more commonly mentioned, although sometimes difficult, ways to increase production is through layout changes. In a research by Centobelli et al. (2016), it was reported that the layout changes along with conveyor system added to a process led to a modest decrease in production time. This study added equipment, making it difficult to determine if the time reduction was truly from the layout changes or from the addition of the conveyor belts.

In another study, Nyati et al, (2017) focused on using systematic layout planning for better streamlining the flow of materials through the plants, making effective utilization of cubic space and maintaining flexibility of arrangement and operation. With the new developed plant layout, the total distance traveled per part was reduced by 50%. Although this article shows a significant reduction in time wasted and manpower required for material handling, no information was given as to how manufacturing throughput was actually affected.

Suthar (2014) also presented a study of increasing productivity by focusing on workflow parameters for machine setup and its process and mentioned changing work cell layout, but no specific improvement statistic was given.

Despite the existing research work on the manufacturing efficiency improvement via layout redesign, its impact as a main factor was not well quantified and how its contribution in conjunction with other factors was not well documented.

PROBLEM STATEMENT & RESEARCH METHODOLOGY

In order to increase throughput, changes need to be made to existing processes. Some of these changes can be made quickly with low risk, such as reallocating employees between work cells. Other changes are high risk and will be very disruptive to production, such as moving entire work cells. The purpose of this study is to quickly test hypothesis, in a simplified manner, to reduce cycle times and increase throughput by modifying the facility layout, as well as reallocating the existing employees between work cells.

All of this takes place in the context of the assembly portion of a hydro excavation truck company's manufacturing process. In addition, the study will look into how the layout redesign and reallocation of employee resources changes combined two changes combined affects the impact the layout redesign makes to improving the production efficiency. Both factors are common changes suggested in the manufacturing industry.

We will adopt the discrete event simulation (using the AnyLogic software) as a main platform to predict the manufacturing system behavior. The 1st simulation model will be based on the current process and will serve as the baseline. The 2nd model will focus on rebalancing employees between work cells with no other changes. Likewise, the 3rd model will focus exclusively on changing the facility layout. The 4th and final model will combine the layout improvements and employee rebalancing improvements. All four models will be compared for percentage changes in truck assembly time. The practical application of this case study will be a presentation of model-based information to management to justify changes to be made in a timely fashion.

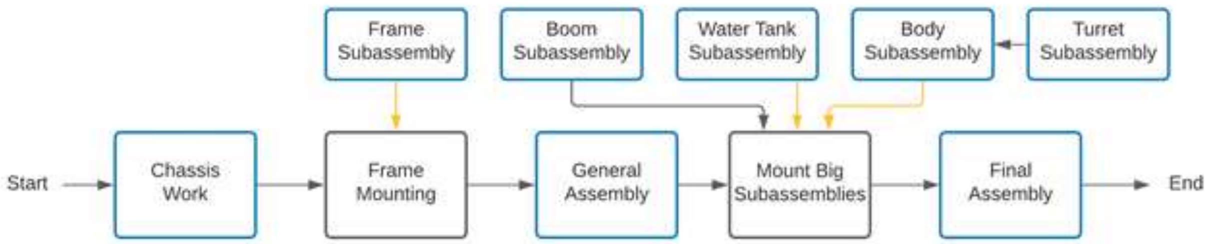


Fig. 1. Process Flowchart

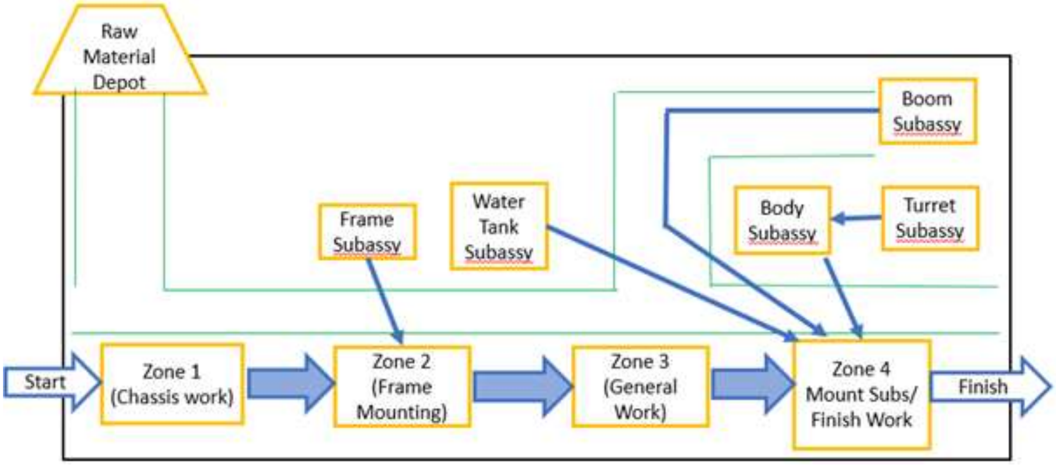


Fig. 2. Facility Layout: flow arrows, work cells (yellow), and aisle ways (green)

SIMULATIONS & OUTCOMES

This research is designed around several discrete event simulation models. Both current and proposed future state models were created for comparison. The current state was designed using the case study's assembly process layout, time studies, and work cell employee numbers. Through best practice layout methods (such as the S, T, and U flow shapes) and simulated cycle times (Johnson, 2012), the proposed future state models were created.

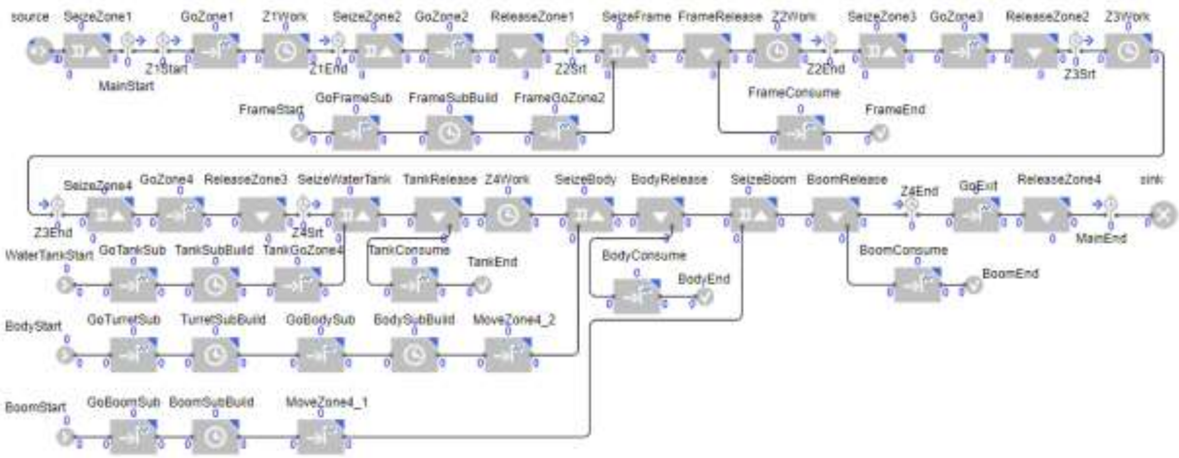


Fig. 3. Base model for the simulation

The model functions on 2 basic algorithms: zone and subassembly. The zone algorithm's function is for the agent (the hydro excavator truck) to move into a zone, delay to simulate work, and move to the next zone when it becomes available. The subassembly algorithm's function is to move to a work area, delay to simulate being built, then move to the zone requesting it. After this, some clean-up logic is used to updating the agent's state while removing the resource from the simulation.

All four models used for the discrete event simulation use a normal distribution (based on actual build data) for hours. The simulation ran for one month's worth of production, which is approximately 20 trucks. The above model logic was used for all 4 simulations, with the only additions being some parameters to change the number of workers in each area and additional timers for models 2 and 4. The work cell movement of models 3 and 4 did not change any of the logic behind the base model.

Outcomes of Base Model

The baseline simulation model was built using data and logistics from the entire assembly process and illustrates the current, non-optimized state and features of the current production. Figure 4 below shows the baseline simulation layout and the current employee allocation.

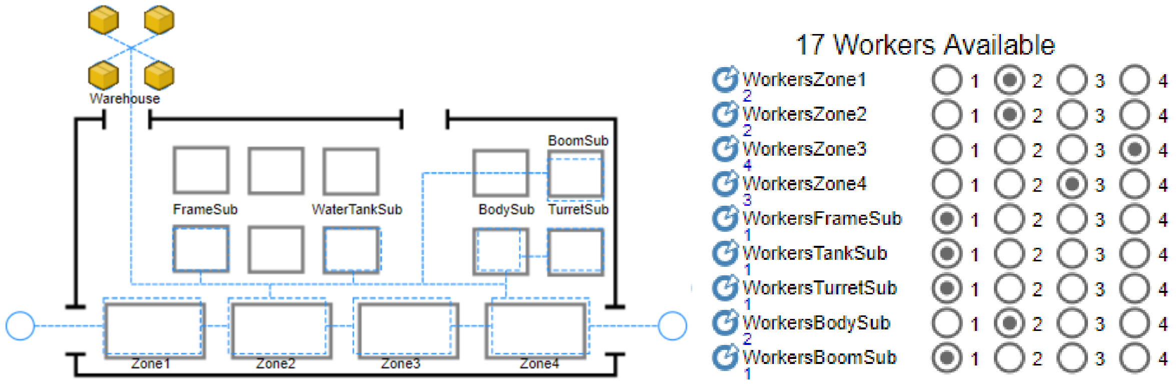


Fig. 4. Base Model 1: Current Layout and Employee Allocation

After running the baseline simulation multiple times, it was discovered that it takes an average of 212 hours with a standard deviation of 26.6 hours to build 20 trucks. We were also able to identify areas in the model which accounted for the most time taken in the process. The zone 4 assembly process accounted for the longest time due to the multiple subassembly processes present in that zone, therefore our main focus in optimizing the employee allocation was to assign as many employees to zone 4 and its subassemblies while also maintaining the other zones and making sure that those processes remain in control.

Outcomes of Model 2

It took multiple tries to find the optimal employee allocation but we were able to come up with an allocation as seen in figure 5 below. Due to the relatively low cycle times in zones 1 and 3, we were able to reallocate employees from those zones to zone 4 and related subassemblies while being constrained to 17 workers. This resulted in a new average of 178 hours with a standard deviation of 17.7 hours to build 20 trucks which is a 16% decrease from the baseline simulation. This is the type of change that management would be looking for.

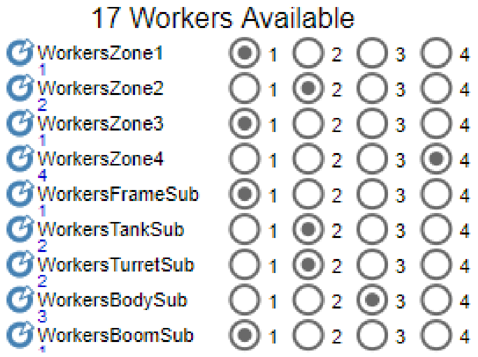


Fig. 5. Model 2: Employee Reallocation Model

Outcomes of Model 3

Model 3 involved rearranging the work cells from the baseline simulation while leaving the employee allocation unchanged. Moving work cells promotes smoother operation in production as well as reducing the movement of resources (which is a form of waste). The new facility layout as presented in figure 6 provides direct paths for movement of resources with the added bonus of safety by reducing fork truck/pedestrian interactions. This resulted in a new average of 210 hours with a standard deviation of 32.1 hours to build 20 trucks, which is only a 1% improvement from the baseline model. This result would tell management to start looking into other options of increasing throughput.

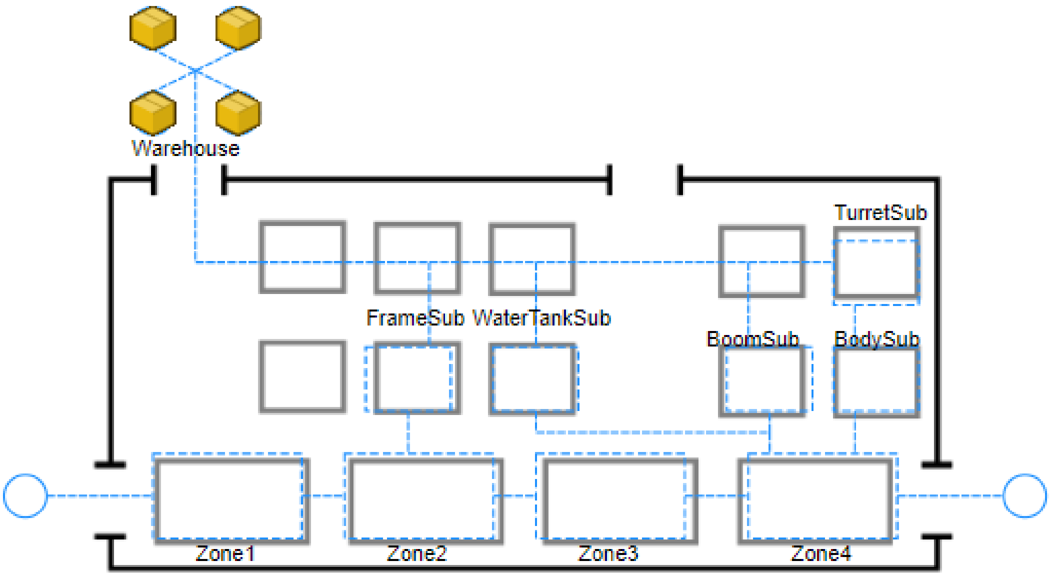


Fig. 6. Model 3: Facility Layout Change

Outcomes of Model 4

This final model combined both model 2 and model 3. The optimal employee allocation and facility layout were combined in order to maximize the effects of both improvements.

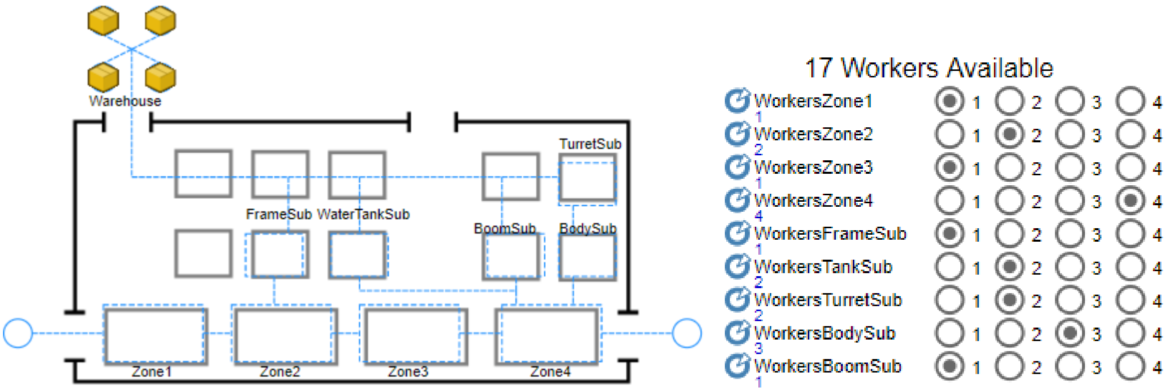


Fig. 7. Model 4: Combination of Employee Reallocation and Facility Layout Change Models

This combined model was able to produce a joint effect that is greater than the sum of the individually achieved effects. The new average time it took to build 20 trucks was 167 hours with a standard deviation of 14.1 hours. This was a 21% improvement from the baseline model which is significant as it reduces the time it takes to build one truck by 45 hours, which allows for about 5 more builds per month. This information can be used to justify both changes occurring concurrently to management.

CONCLUSION

This case study introduced a discrete event simulation-based optimization case for increasing manufacturing throughput through the layout redesigning and employee reallocation. The discrete event method was applied to simulate the operation for re-assigning the employees across different working cells and creating different layouts. The model was used to analyze a typical, benchmarking production plant layout with a comprehensive set of orders and resources, with relatively complex routings and varying, processing times.

The simulation focused on utilizing the layout as the main factor and reallocating available resources as a secondary factor to increase the throughput in manufacturing. While the layout changes alone created an insignificant throughput gain, when it was combined with the employee reallocation, the throughput gain was significant with a 21% throughput increase.

This study showed how a complex and general layout optimization assignment could be done with additional decision factors, to help justify a typical change decision to management. With the outcomes in this study, the management team can utilize the simple simulation techniques in order to preview throughput grains with resources already available to them.

REFERENCES

1. Centobelli, P., Cerchione, R., Murino T., & Gallo, M. (2016). Layout and Material Flow Optimization in Digital Factory. International Journal of Simulation Modelling, 15(2), 1726-4529. http://www.ijsimm.com/Full_Papers/Fulltext2016/text15-2_223-235.pdf

2. DC Velocity Staff. (2017, June 26). Industrial truck industry contributes $25.7 billion to U.S. economy, report says [Web log post]. Retrieved March 30, 2021, from https://www.dcvelocity.com/articles/28966-industrial-truck-industry-contributes-25-7-billion-to-u-s-economy-report-says

3. Gazer, A. (2019, May 28). 4 ways to increase productivity. Retrieved March 22, 2021, from https://getfreepoint.com/4-ways-to-increase-production-capacity-with-the-same-headcount/

4. Gokulraju R, Vigneshwar K, & Vignesh V. (2016). A Case Study on Reducing the Lead Time and Increasing Throughput by using Value Stream Mapping. International Research Journal of Engineering and Technology (IRJET), 3(12). https://www.irjet.net/archives/V3/i12/IRJET-V3I12130.pdf

5. Johnson, I. (2012, August 20). Simplifying Lean Manufacturing: Work Cell Output, Cycle Time Variances & Production Volumes. Retrieved March 22, 2021, from https://www.driveyoursuccess.com/2012/08/simplifying-lean-manufacturing-work-cell-output-cycle-time-variances-production-volumes.html

6. Nyati, V., Jaybhaye, M. D., & Sardar, V. (2017, December 21). Optimization Of Facility Layout For Improvement In Productivity. ResearchGate. https://www.researchgate.net/publication/323358115_OPTIMIZATION_OF_FACILITY_LAYOUT_FOR_IMPROVEMENT_IN_PRODUCTIVITY.

7. Steel, M. (2016, September 28). 6 effective ways to improve Manufacturing throughput. Retrieved March 24, 2021, from https://www.marlinwire.com/blog/improve-manufacturing-throughput

8. Suthar, Kinjal & Deshpande, Vivek. (2014). Review on reducing manufacturing throughput time: Various tools and techniques. 1. 542-546. https://www.researchgate.net/publication/333972193_Review_on_reducing_manufacturing_throughput_time_Various_tools_and_techniques

9. UC Staff. (2019, April 23). Vacuum Truck Market Size To Reach $1.45 Billion By 2024. Retrieved March 30, 2021, from https://utilitycontractoronline.com/vacuum-truck-market-size-to-reach-1-45-billion-by-2024/

10. Vorne. (n.d.). *Increase Manufacturing Production Capacity*. Retrieved March 21, 2021, from https://www.vorne.com/solutions/increase-manufacturing-production-capacity.htm

DECISION SCIENCES INSTITUTE

Artificial Intelligence-based Part Surface Visual Defect Detection

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Pan Zou
Georgia Institute of Technology
Email: zoupzp123@gatech.edu

Xuejian Gong
Georgia Institute of Technology
Email: xgong45@gatech.edu

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Jianyuan Peng
Georgia Institute of Technology
Email: pjy@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

Visual defect detection is an important procedure in production to identify defective products that lead to production yield loss and to implement manufacturing automation. In this paper, a vision-based data analytics and intelligent reasoning system is proposed for mixed-model production lines. Template matching is applied for product classification. The residual neural network (ResNet) is used to detect defects with clear features. Different machine vision techniques are applied for defect feature extraction. Several reasoning tools are utilized to further classify feature data and decide rework operations. An application case of PCB surface defect detection is presented to validate the system.

KEYWORDS: Defect detection, Vision-based data analysis, CNN, Intelligent reasoning, Automatic inspection

INTRODUCTION

In manufacturing industries, quality control is an important procedure during production to guarantee the product quality through defect inspection, thus improving the yield of qualified products and reducing possible manufacturing costs on defective products. According to the production process, defect detection is applied in multiple stages so that defects can be identified in time, such as the layout, fabrication, surface treatment, assembly, and testing processes (Huang & Pan, 2015). For products like automotive and aircraft, destructive testing is sometimes used. While for other manufacturing industries, non-destructive evaluation is often applied. Prevailing non-destructive approaches include the vision-based approach, radiography, ultrasonic testing, the eddy current approach, and thermography (Czimmermann et al., 2020). For example, metal manufacturing uses ultrasonic arrays to measure alloy thickness (Shih et al., 2013), and concrete production applies infrared thermography for embedded defect detection (Cheng et al., 2008). For defects that can be observed on the product surface, the vision-based approach is usually used. Such defects can be categorized into two. The first category is the texture-related defects (Xie, 2008), like scratches and contamination on wood (Phan & Alcock, 1998), steel (Neogi et al., 2014), textile (Ngan et al., 2011), and ceramic tiles (Karimi & Asemani, 2014, Elbehiery et al., 2005). The second category is relating to components attached to the product surface. Such defects normally exist in electronics manufacturing, like semiconductors (Huang & Pan, 2015) and PCBs (Moganti et al., 1996).

Traditional visual defect detection relies on human inspection. However, the inspection accuracy and efficiency are unstable because humans make errors and get tired after continuous and repetitive work. Also, human inspectors need training to learn the identification and classification of different types of defects. Moreover, human inspectors have difficulties detecting tiny defects without the help of a machine. In recent decades, advances in manufacturing technology and theories accelerate the pace of production and facilitate precision manufacturing, presenting a new challenge for product inspection. To seek fast, accurate, and reliable defect detection, human inspection is gradually substituted with the machine vision-based automatic inspection. Therefore, there are growing demands for automatic inspection systems, and research has been on both hardware and algorithm perspectives. This paper mainly discusses the detection algorithms.

Several technical challenges exist in visual defect detection applications, including high-level feature extraction, learning-based model training with a limited defect sample size, and defect identification and classification with extracted features.

(1) High-level feature extraction differs from application to application. Because products are made from different materials, and product defects are defined subjectively in different industries. Thus, high-level features should be uniquely designed to represent different types of defective products and to meet the requirement of the specified application. Furthermore, several factors, like lighting conditions and surface texture, can affect the effectiveness of defect representation with features. Therefore, selected features should characterize the differences between good products and defective products in different scenarios.

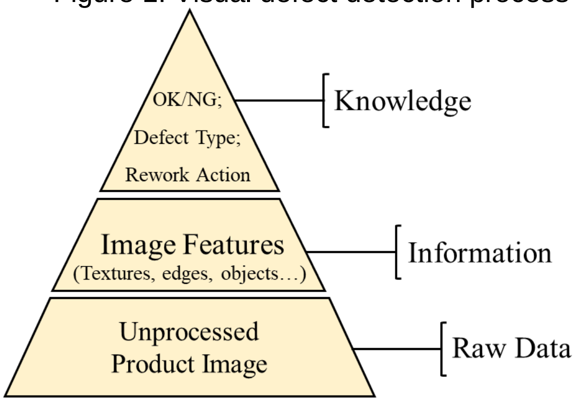
(2) The training of learning-based approaches usually requires a large data sample size. The model robustness and the generalization ability of convolutional neural network (CNN) models are largely determined by the number of training data. The requirement for the number and the quality of data also grows higher as more defect types are defined. However, useful data usually has a limited volume in practice (Deng et al., 2018), which makes model training a challenging task. Also, deep learning-based models have difficulties learning defect features when defects are irregular and small in shape when the data sample size is small.

(3) Defect identification and classification require fast and reliable detection to understand product quality. Industries make great efforts to improve production rates, and the manufacturing activities are conducted in real-time. Thus, there is a trade-off between the time and the accuracy of defect detection. Because feature extraction takes time, and the number of extractable features is not limited, choosing a subset of features that meet the requirement for both time and accuracy is necessary. Moreover, extracted features can be either numeric or nominal data. Therefore, proper

classification algorithms should be selected to process these two different data types. Furthermore, some types of defects are classified based on experience instead of clearly stated rules. In this case, algorithms should consider such vagueness during defect identification and classification.

Even though the defect feature extractors should be designed based on the specific application, the detection process can be generalized as two tasks from the information perspective, as is shown in Figure 1. The first task is to obtain image information from raw 2D images. This can be seen as visual data analysis through visual processing techniques. The second task is to derive knowledge from image information. This usually involves a series of decision making through reasoning and predictions. The knowledge is about application objectives: the detection result, the defect type, and the rework operations.

Figure 1: Visual defect detection process



Prevailing techniques for visual defect detection include visual processing, deep learning, and symbolic reasoning, and each technique has its strengths and weaknesses. For example, visual processing operates on the raw image to obtain desired features and does not require a large dataset for algorithm robustness, but features should be designed manually. The deep learning-based methods implement automatic feature extraction and classification using raw images, but much high-quality data is needed for model training to guarantee the generalization ability. Symbolic reasoning can address the vagueness during defect identification and classification by mimicking the human reasoning process, but the performance depends on the selection of defect features.

In practice, products from different product families are manufactured simultaneously in one production line, which requires the inspection system to classify different product families and to detect defects on different product surfaces. Meanwhile, the data volume is sometimes insufficient to train a single model for all the tasks. In this regard, this paper proposes a vision-based data analytics and intelligent reasoning system that combines the strengths of different techniques to address the discussed technical challenges. Vision-based data analytics includes visual processing algorithms, CNNs, and classification algorithms, while intelligent reasoning involves rough set theory (RST) and case-based reasoning (CBR). In this system, product classification is first implemented by feature extraction. The CNN model is developed to identify noticeable defects. Data augmentation and transfer learning are applied to lower the requirement for the defect sample size. For smaller defects that are difficult to identify, visual processing techniques are applied to extract features for further analysis. If the feature data is numeric, classification algorithms like support vector machine (SVM) are used to develop the classification model. If the feature data contains nominal information, and much vagueness exists during decision making, RST can be utilized. Finally, the CBR model evaluates the defective products based on extracted features to assign corresponding rework operations.

The rest of this paper is organized as follows. The background and related work for visual defect detection is discussed. Then the analysis and design of the proposed vision-based data analytics and intelligent reasoning system are presented. The visual processing techniques for feature extraction are introduced. The CNN-based defect identification is discussed, and an improved ResNet is presented. The application of reasoning techniques, including RST for defect identification and CBR for rework operation assignment, and corresponding knowledge representation are presented. An application case of defect detection during PCB surface treatment is presented. Finally, the application case and potential benefits in manufacturing industries are discussed.

LITERATURE REVIEW

Visual Defect Detection

Visual defect detection aims to identify and classify defective products based on predefined rules. The process starts from data acquisition, where the camera system obtains the product images from the production environment. Image preprocessing operations, like noise removing and segmentation, may be needed for data preparation. Image features are extracted to describe the product in artificial representation using visual processing techniques. Because the classification is a process to map the feature data to the defect type, it can be modeled as a series of decisions guided by predefined rules, or the mapping relationship can be learned mathematically. Therefore, defect identification and classification are implemented through reasoning-based approaches (Chang et al., 2008) or learning-based approaches (Zhang et al., 2018; Czimmermann et al., 2020). After determining the defect type, rework operations can be assigned according to the feature data.

Visual defects can be categorized into texture-related defects and component-related defects. Some defects occur on the surface of the product and cause visible texture abnormality. The prevailing taxonomy of such defects can be referred to (Czimmermann et al., 2020), in which palpable defects sometimes need manual touch. Texture analysis is usually applied for the detection of these defects.

The other group of defects is caused by the components attached to the product surface, and this is commonly seen in electronics products. Examples of component-related defects include misplacement, missing, over etch/under etch, wrong orientation, and wrong size. Because the number of components on products can be large, and they have different shapes, algorithms that measure similarities like template matching and image subtraction (Chang et al., 2008) are sometimes very effective.

Visual Processing

Based on the defect category, different visual processing techniques are applied. If the product has texture-related defects, like textiles and steel, texture analysis is implemented (Xie, 2008; Czimmermann et al., 2020). For products with components attached to the surface and components are the cause of defects, referential algorithms that use a faultless image for comparison, like template matching, are applied (Hani et al., 2012).

Common approaches for texture analysis involve statistical approaches, structural approaches, and filter-based approaches. Statistical approaches implement statistical analysis on the image pixel matrix directly. The analysis includes the histogram analysis (Ng, 2006), principal component analysis and independent component analysis (Chen & Perng, 2011; Serdaroglu et al., 2006), gray level co-occurrence matrices (Rahja et al., 2013), auto-correlation (Zhu et al., 2015), and local binary pattern (Mäenpää & Pietikäinen, 2005). Structural approaches extract structural elements, like edges and corners, and analyze the spatial arrangement of them. Standard methods are edge and skeleton representation (Wen & Xia, 1999, Chen & Jain, 1988) and morphological operations (Mallik-Goswami & Datta, 2000). Filter-based approaches analyze texture information from not only the spatial domain, but the frequency domain and the joint spatial/spatial-frequency domain. Well-known methods include Gabor filtering (Kumar & Pang, 2000) and Wavelet filtering (Lambert & Bock, 1997).

Images of products with mounted components contain more details in surface objects, and texture information is less important. Usually, the target is compared with defect-free images for defect detection, and typical methods are template matching and image subtraction. Template matching can measure the similarity between the target image and the reference image, which should work for component misplacement, missing, misaligned, and rotated components (Crispin & Rankov, 2007). Image subtraction compares two images with the XOR operator pixel by pixel, and defects are detected based on the difference matrix. This method requires the image to be aligned and binarized first, and sometimes segmentation is also implemented to locate and label different regions of interest (Chang, 2008, Gonzalez, 2002).

Symbolic Reasoning

Symbolic reasoning mimics the human decision-making process and can address the vagueness during defect detection, which is especially helpful when the classification is based on experience. Prevailing approaches include rough set theory, fuzzy logic, rule-based reasoning, and case-based reasoning. Fuzzy set theory is a successful theoretical approach to vagueness (Zbigniew, 2004), and fuzzy inductive reasoning can be implemented on image histogram statistics (Ye, 2009). Rough set theory is another mathematical approach to vagueness and imperfect knowledge (Pawlak, 1998). It is helpful when there are nominal factors in extracted features. Compared to other knowledge-based systems, CBR does not require much effort in a domain knowledge model but solves problems by adapting previously successful solutions to similar problems (Watson & Marir, 1994). An application example is to determine the possibility of PCB circuit defects by comparing defect features with past defect cases (Chang et al., 2008).

Deep Learning for Defect Detection

Deep learning is a subclass of machine learning techniques, which learns the representation of high-level features to represent classification objectives. In defect detection, the main applications are in image classification and object detection. Image classification refers to labeling an image with a probability that it falls into a predefined object class. Data can be labeled as defect-free products and products with different types of defects for defect classification. An example is the use of a five-layer CNN to detect 6 types of defects (Weimer et al., 2016). While object detection can detect the presence of a group of objects in an image, with the probability and position estimation of each instance (Guo et al., 2016). And this implements defect classification and defect localization simultaneously. An example is to use faster R-CNN with sliding windows for a surface inspection system (Wang et al., 2020). Conventional CNN models for image classification are VGG (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016 a). And region-based CNNs like Mask R-CNN are prevailing approaches for object detection (He et al., 2017).

ANALYSIS AND DESIGN OF THE VISION-BASED DATA ANALYTICS AND INTELLIGENT REASONING SYSTEM

To address the vagueness during defect identification and classification and the difficulties caused by the limited data sample size, this paper proposes a vision-based data analytics and intelligent reasoning system using combined techniques. This section presents the system functional analysis.

The proposed system implements image data analysis and reasoning for defect detection and uses hybrid techniques, including visual processing, learning-based classification algorithms, CNNs, and symbolic reasoning. The functional analysis is illustrated below. Firstly, image data is obtained and stored as a 2D matrix in the computer from the calibrated camera system. Following a fixed sequence, image preprocessing implements several visual processing algorithms to clean up the data and to prepare for further operations. Then product structural features are extracted, and template matching is implemented for product classification. The CNN model in this system is trained to identify defects that can be easily recognized. Thus, the training requires fewer data since the features are not complex. In this case, the OK dataset is constructed with real OK data and defect data that needs further analysis. If the CNN classifies the product as defective, the system directly outputs the defect type. Otherwise, further defect features are extracted through morphological operations or texture analysis. The feature vector of the inspected product is created for defect identification and classification. If there exist clearly stated rules to classify defect types, a rule-based model can be developed. Otherwise, classification algorithms like SVM can be utilized to model the relationship between feature data and defect types. And if vagueness exists during decision making, or much nominal feature data is collected, RST will be applied to extract classification rules. If the product is defective, the feature data is analyzed with the defect type by CBR to assign a proper rework operation.

VISUAL PROCESSING FOR FEATURE EXTRACTION

This section discusses visual processing techniques for feature extraction, including image preprocessing algorithms, morphological processing, and texture analysis approaches.

Image Preprocessing

The objective of image preprocessing is to clean up and prepare data for further feature extraction. Common preprocessing operations include noise reduction, image enhancement, and segmentation.

Noise reduction is the process of removing noise from the image signal. While smoothing the image, it also blurs the details of small features in the image. Prevailing noise reduction algorithms are mean filtering and Gaussian filtering.

Image enhancement is a series of adjustments in image brightness, contrast, saturation, and so on. In grayscale images, contrast enhancement and binarization are two common operations. Histogram equalization is a direct way to enlarge the contrast. And binarization is to convert the multi-level image matrix into a two-level matrix by choosing a proper threshold. The pixel larger than the threshold is turned into white, and the pixel smaller is turned into black. Conventional thresholding methods are Otsu's thresholding and adaptive thresholding (Otsu, 1979; Bradley & Roth, 2007).

Segmentation is to separate the background from the foreground objects and thus segment the image into different regions. The objective is to simplify image representation so that the image is more meaningful and easier to analyze. Four approaches to segmentation are thresholding, clustering methods, compression-based methods, and histogram-based methods.

Morphological Processing

Morphological processing is to process images based on shapes. It uses a structuring element smaller than the image and implements some basic operations to generate a new image. Four common operations are erosion, dilation, opening, and closing.

Morphological processing operates differently in binary images and grayscale images. In a binary image A, the dilation operation with a structuring element B is defined as follows:

$$A \oplus B = \left\{ x \middle| \left(\hat{B}\right)_x \cap A \neq \emptyset \right\} \tag{1}$$

Thus, the white region grows larger after the dilation operation. In contrast, the erosion operation makes the white region smaller, and it is defined as follows:

$$A \ominus B = \{x|(B)_x \subseteq A\} \tag{2}$$

The opening operation is to implement erosion followed by dilation. And closing is to implement dilation followed by erosion.

In grayscale images, erosion and dilation are both pixel by pixel operations. Erosion is to replace the current pixel with the minimum pixel value within the neighbor region B, while dilation is to replace it with the maximum pixel value. The definition of opening and closing is the same as that in the binary image. Thus, erosion is often used to reduce the size of bright features and increase the size of dark features, while dilation does the opposite. Opening is used to remove small and bright details but leave the overall intensity level and larger bright features unchanged. And closing has the opposite effect.

Texture Analysis

Prevailing texture features can be analyzed in three approaches: the statistical approach, the structural approach, and the filter-based approach (Czimmermann et al., 2020).

The statistical approach is to implement statistical analysis of the image pixel matrix, and it is useful when structural features are unable to describe the defect. Prevailing statistical methods include the histogram statistics, the co-occurrence matrix, the autocorrelation, and the local binary pattern. The image histogram contains the first-order statistical features. Its implementation is computationally cheap. The co-occurrence matrix, which counts the occurrence of pixel value pairs, indicates much information about texture features (Conners et al., 1983). Several measures are defined from the co-occurrence matrix. Four different orientations are usually chosen to obtain rotation-invariant measures. The autocorrelation method is to measure the correlation between the texture image and its transformation with different displacements to detect repetitive feature patterns. And local binary pattern is an illumination-invariant and rotation-invariant method to extract texture edge features with low computational costs.

The structural approach aims to extract geometric information from the original image. Commonly used features are edges, corners, curves, shapes, and the skeleton. Corresponding algorithms are Canny edge detection, Harris corner detection, the Hough method, and the curvature-based method.

The filter-based approach implements frequency domain analysis or joint spatial/spatial frequency analysis. The objective is to analyze the image components and process them in frequency domain, which proves efficient in noise reduction. A popular filter for texture feature extraction is the Gabor filter, and it can segment components of different shapes by using a set of filters. Thus, texture features of interest can be extracted separately for further analysis.

DEFECT IDENTIFICATION USING THE IMPROVED RESNET WITH LIMITED TRAINING DATA

The CNN is a deep learning model with multiple layers trained in a robust manner. It is found to be highly effective in learning high-level feature representation in the image processing field. However, several limitations should be considered about using CNNs for defect detection. Firstly, the number of data samples in different classes should be similar to avoid bias. Secondly, feature learning can be difficult if the defect is too small. Thirdly, CNNs cannot adjust the model specificity. In this section, basic CNN layers are introduced. Data augmentation and transfer learning are discussed as the training strategies when the data sample size is small. And an improved ResNet is introduced as the defect identification model.

Model Training with Data Augmentation and Transfer Learning

Data augmentation and transfer learning are two training strategies when the dataset is small. The former enlarges the original dataset, while the latter lowers the requirement of the dataset size.

Because of the translation invariance of convolution, geometric transformations can be done on original images, thus generating more data to improve the model generalization performance. Common transformations include flipping, cropping, rotation, translation, and noise injection. Sharpening images with kernel filters is also helpful (Shorten & Khoshgoftaar, 2019). During the model training, different augmentation operation combination is normally taken. However, two issues are worth noticing. The first is that massively inflating the dataset may result in further overfitting when the original dataset is small. The second is the safety of the augmentation operation. The transformation should not alter the label of the original image. Otherwise, the training will be done with wrongly labeled images.

Transfer learning is to initiate the network parameters with a pre-trained model because the first several convolutional blocks generate the general features that can be used for other classification tasks. Thus, the parameters in these layers can be frozen, and only parameters in the classification layer and the last several layers should be trained. In this way, the training time is largely saved, and less data is needed to develop a new classification model.

An Improved Residual Neural Network Model

The improved ResNet-50 model is chosen as the model for defect identification. Compared to other models like GoogleNet and VGG16, ResNet is known for its generalization performance and computation cost. It is also deeper than VGG16 but has fewer network parameters, which makes it faster than the VGG16. One challenge the ResNet solves is that the training error grows higher when the network is deeper and deeper, which is caused by the vanishing gradient problem. And ResNet addresses this problem by its skipping connections: making a layer of inputs skip some intermediate layers and jump directly to the end of the residual block (He et al., 2016a). This makes each few stacked layers fit a residual mapping instead of fitting a desired underlying mapping.

The improved ResNet changes the residual block to improve information flow through the network. Compared to the original block, batch normalization and the ReLU activation function are before 2D convolution (He et al., 2016b). This change further eases the training process and improves generalization.

SYMBOLIC REASONING FOR DEFECT DETECTION

This section introduces reasoning techniques for defect detection. Compared to learning-based classification algorithms, symbolic reasoning is similar to the human decision-making process, which is an advantage to deal with vague information. In this study, the rough set theory and case-based reasoning are applied, in which the former is applied for defect classification and the latter for rework operation assignment.

Rough Set Theory for Defect Classification

Similar to SVM, RST can be applied as a classifier to process the feature vector for defect detection by extracting classification rules. Different from numerical analysis, RST studies the relationship between the conditional attributes and the decision attribute in the form of the production rules (IF-THEN), and it expresses the imprecision or vagueness in the decision system by the boundary region of a set. The result of RST is a set of conditional rules that indicate the map between conditional attributes to the decision attribute.

The implementation of RST has several steps: data discretization, attribute reduction, the study of indiscernibility relation, and rule induction. Data discretization is to divide the numeric data into different regions so that numeric data can be converted into nominal data. Conventional methods for data discretization include the global discernibility algorithm, which computes the globally semi-optimal cuts using the maximum discernibility heuristic, the quantile-based discretization, and discretization by equal intervals. The attribute reduction is to find the reducts of current conditional attributes from the discernibility matrix. The rejected attributes should be redundant ones whose removal will not worse the classification. The common approach for attribute reduction is the heuristics, while the reduct generation is either based on different criteria, like entropy and discernibility measure, or on a permutation schema over all attributes. The indiscernibility relation refers to a subset of attributes by which two objects are indiscernible. IF-THEN decision rules can then be derived from the indiscernibility classes defined by a subset of attributes.

Rework Operation Assignment with Case-based Reasoning

After determining the defect type, case-based reasoning is applied to determine the rework operation for the defective product. The advantage of CBR over pure production rules is that CBR does not require much domain knowledge to make a choice. Instead, it relies on the accumulation of solved cases and finds the solution to its own by comparing the difference of the feature vectors. This makes the CBR system more robust and evolutionary as new cases are continuously put in. Generally, the implementation of CBT involves the following steps: case representation and indexing, new case creation, case retrieval, case reuse or case adaptation, and case retention.

Case representation and indexing is the preparation for further steps. Since CBR relies on knowledge sharing from past cases, a case library should be developed to keep the solved ones. Because the objective is to determine the rework operation for a defective product, the location of the product image, the defect type, its feature vector, and its rework operation should be recorded. Like a database, each case should be indexed properly for reference efficiency. After the development of the case library, a new defective product will be encoded as a new case to reason its solution, whose initialization follows the format of case representation. Then case retrieval is conducted to search for similar defect cases in the case library. This process is also related to how a case is indexed. The similarity should be measured with each case in the same defect class to find the most similar one using a specified similarity measuring algorithm. Based on the similarity value and available rework operations for the current defect type, case reuse or case adaptation can be done to find the solution to the new case. Case adaptation usually involves IF-THEN rules that indicate how the case should be adapted based on the difference between the two feature vectors. After the solution is found, the solved new case can be stored in the case library for future reference.

AN APPLICATION CASE OF DEFECT DETECTION DURING PCB SURFACE TREATMENT

This section presents a case of PCB defect detection during surface treatment. PCB surface treatment is a process of artificially forming a surface layer on the PCB. In this case, the detection region is the gold finger area. This area is to connect multiple boards and consists of several gold-plated connectors, which are horizontally placed side by side, and the quality of these connectors determines the conductivity (PCB Directory, 2020). Defective connectors have contamination spots on their surface. Thus, the inspection mainly focuses on texture information and aims at detecting spots that affect the product quality or result in production yield loss (Zhang & Luk, 2007). The rest of this section presents the implementation details of the vision-based data analytics and intelligent reasoning system.

Problem Description

In this case, gold finger defect detection is implemented for PCBs from two product families. The inspection objective is to detect two types of contamination spots on the gold finger surface: the defective board either has one or two large spots or many tiny spots in some region. The shape of both spots is not a circle but irregular. In a 200*700 gold finger image, a large spot can take up to 40*40 pixels, while a tiny spot occupies less than 5*5 pixels. In some data, normal texture that has different pixel values from the background and randomly exists on the surface, making the detection of tiny spots difficult. Therefore, the identification of defective PCBs with tiny spots is based on experience, and there are no clear rules for classification. The defect sample size for each product family and each side is limited: 150 images with large spots, 120 images with tiny spots, and 40 images with no defects.

The workflow is shown in Figure 2. Firstly, several image preprocessing operations, like Gaussian smoothing and morphological operations, are conducted on the PCB image. Then the structural features of the PCB are extracted, based on which template matching is implemented for product family and PCB side classification. After determining the product type, the gold finger region is cropped out, and the background is removed for defect detection. Considering the number of defective data and defect-free data is small and unbalanced, and the features of tiny spots are challenging to extract, an improved ResNet is trained to identify those with large spots first. The gold finger is then divided into several subregions, and only one connector is kept in each region. Texture analysis is implemented to obtain the feature vector of each region. Then RST or SVM is used to identify regions with tiny spots. Thus, tiny spot identification and localization are both completed. Finally, the system outputs the result for human inspectors as a reference.

Figure 2: The workflow of PCB gold finger defect detection

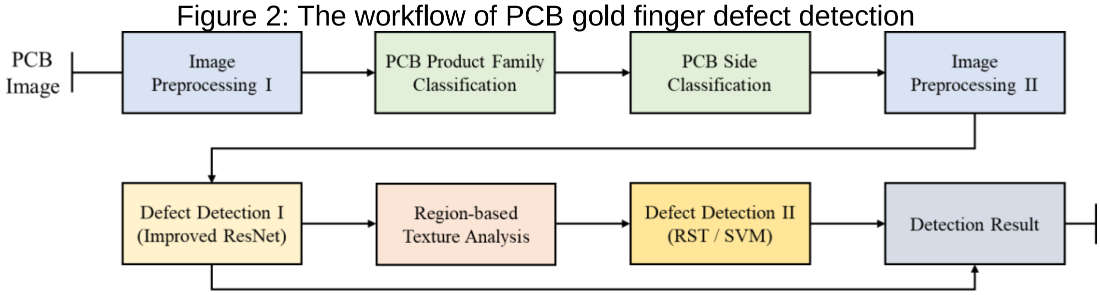


Image Preprocessing and Product Classification

Preprocessing is implemented twice in this case. The first time is for product classification. The reason for product classification is to develop a defect detection model for each type of product so that fewer data will be needed compared to developing one model for all the products. For PCBs, two sides of the same product are different. Thus, both product family classification and side classification are needed. Usually, PCBs from two different product families have many distinguishable structural features, and they can be classified without feature extraction using template matching. In contrast, two sides of a PCB share many structural features. Thus, the objective of image preprocessing is to prepare for the extraction of distinguishable features on two sides. Gaussian smoothing is applied first to remove the noise. Histogram equalization is then used to enlarge the contrast, so the structural features can be more easily extracted. Then morphological operations are conducted on the grayscale image to remove unnecessary details and enlarge structural features of some key components, like chips and resistors.

The second image preprocessing is for defect detection in the gold finger area, and segmentation is the main objective. Histogram equalization is firstly conducted on the original PCB image. The gold finger area is cropped out using a set of predefined ratios corresponding to the product family. Erosion is then implemented to enlarge the size of dark spots. Then Otsu thresholding is done to segment the connectors from the background. Thus, the input into the defect identification model only keeps the details of connectors.

For product family classification, the two sides of defect-free images are concatenated horizontally as templates. Each target image is matched with templates of different product families to determine its classification. For side classification, the edge feature is extracted from the preprocessed image using adaptive thresholding, and the generated binary image is used as the template. Two templates are used to represent the pattern of the two sides. The target image will be processed in the same way before the template matching. The experiment results show accuracy for product family classification and PCB side classification reach 100% with cycle time around 0.1 second.

Large Spot Detection with the Improved ResNet50

There are several reasons not to use CNN to detect tiny spots. The first reason is the dataset composition. The defect-free data is far less than the defective data, and there will be a large bias if the defect-free PCB is defined as one class. Also, the dataset is inefficient to train a CNN with a high generalization ability. The second reason is about tiny spots. This type of defect does not have a regular shape, and its dimension is also very small, making it difficult for CNNs to learn its feature representation. Thirdly, the classification of tiny spot defects is based on experience, and there are no clear rules for the classification. Some PCBs carrying few tiny spots are recognized as defect-free. This brings problems to data labeling and will influence the classification accuracy. Lastly, it is difficult to adjust the specificity of a CNN model, which is an important KPI for abnormality detection tasks. Therefore, the dataset is labeled with two classes: PCBs with large defects and PCBs with tiny defects as well as defect-free PCBs. And the dataset consists of the gold finger regions after background removal, and it is split into three: 60% training data, 20% validation data, and 20% test data.

The improved ResNet50 is chosen as the classification model. Data augmentation and transfer learning are applied to address the problem brought about by the sample data size. Rotation, flip, horizontal shift, and vertical shift are used as augmentation operations. The range of the shift operations is set as 8% to assure the safety of the transformation. The pretrained model is trained on the ImageNet dataset, with the input shape 224*224*3. The output layer of the model is changed, and the activation function is set as the Sigmoid function. The parameters of the other layers are frozen.

The training accuracy of the best model reaches 99.81% accuracy on the training dataset and 100% accuracy on the validation dataset, while the model still has 100% accuracy on the testing dataset. The prediction time is 0.13 seconds on Google Colab GPU. Thus, it can be concluded that the CNN model is effective in identifying large spot defects in the gold finger area.

The training accuracy of the best model reaches 99.81% accuracy on the training dataset and 100% accuracy on the validation dataset, while the model still has 100% accuracy on the testing dataset. The prediction time is 0.13 seconds on Google Colab GPU. Thus, it can be concluded that the CNN model is effective in identifying large spot defects in the gold finger area.

Region-based Texture Analysis on Gold Finger Regions

Leaving out the large spot defect, the rest task is to classify the defect-free products and the ones with tiny spots. The main difficulties of tiny spot detection are as follows. Firstly, there is normal texture that has different pixel values from the connector background and is larger than tiny spots. Its existence influences the detection of tiny spots. Secondly, lighting conditions can be uneven in the horizontal direction. Because connectors are arranged side by side, some regions can be dark while some others are bright. Thirdly, the shape of the defects is not a circle but irregular. Thus, circle detection is not useful in this case.

To overcome the above difficulties, a Gabor filter bank is tested trying to remove the normal texture from the frequency domain while keeping the tiny spot defects. The filter with the best result is chosen. Then the gold finger region is divided into multiple subregions so that there is only one connector in each subregion. This makes the lighting condition relatively fixed locally. As a result, defect detection is implemented on each subregion, and detection results will indicate the defective regions, which serve as a reference for human inspectors.

To extract the texture features, statistical analysis is done on the global image and local subregions. Firstly, the histogram statistics are collected for the gold finger region and all the subregions, including the mean, standard deviation, skewness, and kurtosis of the histogram. Secondly, the co-occurrence matrix of each subregion is obtained, and four features are collected: contrast, homogeneity, energy, and correlation. Thus, the feature vector consists of twelve description attributes.

Tiny Spot Identification Using RST and SVM

After obtaining the feature vector for each subregion, tiny spot identification is to classify it based on the vector. The classification can be either done through intelligent reasoning or mathematical predicting, depending on the data characteristics and the vagueness during defect detection. In this study, SVM and RST are used on two datasets constructed with images from two product families. The product family A dataset has less normal textures, and the Gabor filter can remove most of the normal texture. Also, tiny spots are more densely populated on the connector surface. The product family B dataset has many normal textures, which the Gabor filter can hardly remove. Meanwhile, some data has a limited number of tiny spots, making classification ambiguous and more based on experience. Thus, much vagueness exists in the second group during defect detection.

Both SVM and RST are applied to the two datasets for comparison of detection accuracy in different scenarios. Seven hundred records of product family A are used, in which 564 are for training and 136 for testing. Eight hundred records of product family B are used, of which 640 are for model training or rule induction and 160 for validation. Exhaustive experiments are done to find the feature combination that has the highest detection accuracy first. The experiment results are shown in Table 1.

Table 1: Experiment Results of Tiny Spot Detection Using RST and SVM

	Method	Accuracy	Time (per PCB image) (Feature extraction included)
Product Family A	RST	85.4%	0.75s
	SVM	93.4%	0.80s
Product Family B	RST	75.6%	0.93s
	SVM	73.6%	0.97s

It should be noticed that the detection accuracy is for a single subregion, and the detection of different regions is independent of each other. Considering a PCB has 19 subregions, plus the defective PCBs usually have more than 3 subregions with tiny spots, the probability of not detecting a defective product would be very small. Because the product family B dataset does not have a high quality, the detection accuracy is not satisfying. Still, it is found that RST performs better when vagueness is involved or when the classification is sometimes ambiguous. And SVM has a better performance when defective products can be identified with certainty.

DISCUSSIONS

The application case presents visual defect detection on a mixed-product production line. Both the data quantity and quality are limited, and vagueness exists during tiny spot detection. The system firstly attracts structural features for product classification with template matching, and it uses the CNN model to detect large spots on the gold finger area, which are relatively easy to recognize. Then the gold finger region is cropped out and divided into 19 regions for further texture analysis. Features of tiny spots are extracted from the image histogram and the co-occurrence matrix. Both RST and SVM are used to validate their effectiveness in different scenarios. For each product image, the whole detection process is within 2 seconds, which is smaller than the cycle time in practice.

Potential application fields of the proposed system include the textile industry, the wood industry, the ceramic tile industry, the steel industry, the stone industry, and the electronics industry. It helps manufacturers in mechanized mass-production industries to reduce labor costs, improve production efficiency, and facilitate manufacturing automation. Compared to the traditional machine vision system, design and development of the proposed inspection system will less depend on highly skilled machine vision technicians, because CNN-based defect detection implements automatic feature design and extraction. In the meantime, required testing and development resources can also be reduced, because most of the CNN algorithms are open to the public and developed CNN models can be reused on different projects.

CONCLUSIONS

This paper proposes a vision-based data analytics and intelligent reasoning system for visual defect detection in grayscale images. Apart from modeling the detection process with technical details, the system also aims to enhance the reasoning and learning ability to address the vagueness during defect classification and lower the requirement of the data sample size by combining the strengths of different techniques. The application case of PCB defect detection during surface treatment is presented to validate the feasibility and potential of the proposed system. The case also presents a region-based detection method to deal with tiny defects with irregular shapes and uneven lighting conditions in the inspection region. Future work of this study will focus on defect feature selection and vagueness modeling using fuzzy rough set theory during defect detection.

REFERENCES

Bradley, D., & Roth, G. (2007). Adaptive thresholding using the integral image. Journal of graphics tools, 12(2), 13-21.

Chang, P. C., Chen, L. Y., & Fan, C. Y. (2008). A case-based evolutionary model for defect classification of printed circuit board images. Journal of Intelligent Manufacturing, 19(2), 203-214.

Chen, S. H., & Perng, D. B. (2011). Directional textures auto-inspection using principal component analysis. The International Journal of Advanced Manufacturing Technology, 55(9-12), 1099-1110.

Cheng, C. C., Cheng, T. M., & Chiang, C. H. (2008). Defect detection of concrete structures using both infrared thermography and elastic waves. Automation in Construction, 18(1), 87-92.

Conners, R. W., Mcmillin, C. W., Lin, K., & Vasquez-Espinosa, R. E. (1983). Identifying and locating surface defects in wood: Part of an automated lumber processing system. IEEE Transactions on Pattern Analysis and Machine Intelligence, (6), 573-583.

Crispin, A. J., & Rankov, V. (2007). Automated inspection of PCB components using a genetic algorithm template-matching approach. The International Journal of Advanced Manufacturing Technology, 35(3-4), 293-300.

Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C. M., & Dario, P. (2020). Visual-Based Defect Detection and Classification Approaches for Industrial Applications—A SURVEY. Sensors, 20(5), 1459.

Deng, Y. S., Luo, A. C., & Dai, M. J. (2018, September). Building an automatic defect verification system using deep neural network for pcb defect classification. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 145-149). IEEE.

Elbehiery, H., Hefnawy, A., & Elewa, M. (2005). Surface defects detection for ceramic tiles using image processing and morphological techniques.

Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27-48.

Hani, A. F. M., Malik, A. S., Kamil, R., & Thong, C. M. (2012, January). A review of SMD-PCB defects and detection algorithms. In Fourth International Conference on Machine Vision (ICMV 2011): Computer Vision and Image Analysis; Pattern Recognition and Basic Technologies (Vol. 8350, p. 83501P). International Society for Optics and Photonics.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In European conference on computer vision (pp. 630-645). Springer, Cham.

Huang, S. H., & Pan, Y. C. (2015). Automated visual inspection in the semiconductor industry: A survey. Computers in industry, 66, 1-10.

Karimi, M. H., & Asemani, D. (2014). Surface defect detection in tiling Industries using digital image processing methods: Analysis and evaluation. ISA transactions, 53(3), 834-844.

Kumar, A., & Pang, G. K. (2000). Fabric defect segmentation using multichannel blob detectors. Optical Engineering, 39(12), 3176-3191.

Lambert, G., & Bock, F. (1997, October). Wavelet methods for texture defect detection. In Proceedings of International Conference on Image Processing (Vol. 3, pp. 201-204). IEEE.

Mäenpää, T., & Pietikäinen, M. (2005). Texture analysis with local binary patterns. In Handbook of pattern recognition and computer vision (pp. 197-216).

Mallik-Goswami, B., & Datta, A. K. (2000). Detecting defects in fabric with laser-based morphological image processing. Textile Research Journal, 70(9), 758-762.

Moganti, M., Ercal, F., Dagli, C. H., & Tsunekawa, S. (1996). Automatic PCB inspection algorithms: a survey. Computer vision and image understanding, 63(2), 287-313.

Ng, H. F. (2006). Automatic thresholding for defect detection. Pattern recognition letters, 27(14), 1644-1649.

Ngan, H. Y., Pang, G. K., & Yung, N. H. (2011). Automated fabric defect detection—a review. Image and vision computing, 29(7), 442-458.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

Pawlak, Z. (1998). Rough set theory and its applications to data analysis. Cybernetics & Systems, 29(7), 661-688.

PCB Directory 2020, accessed September 1 2020. https://www.pcbdirectory.com/community/what-are-pcb-gold-fingers.

Phan, D. T., & Alcock, R. J. (1998). Automated grading and defect detection: a review. Forest Products Journal, 48(4), 34.

Raheja, J. L., Kumar, S., & Chaudhary, A. (2013). Fabric defect detection based on GLCM and Gabor filter: A comparison. Optik, 124(23), 6469-6474.

Serdaroglu, A., Ertuzun, A., & Erçil, A. (2006). Defect detection in textile fabric images using wavelet transforms and independent component analysis. Pattern Recognition and Image Analysis, 16(1), 61-64.

Shih, J. L., Wu, K. T., Jen, C. K., Chiu, C. H., Tzeng, J. C., & Liaw, J. W. (2013). Applications of flexible ultrasonic transducer array for defect detection at 150 C. Sensors, 13(1), 975-983.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

Wang, Y., Liu, M., Zheng, P., Yang, H., & Zou, J. (2020). A smart surface inspection system using faster R-CNN in cloud-edge computing environment. Advanced Engineering Informatics, 43, 101037.

Watson, I., & Marir, F. (1994). Case-based reasoning: A review. Knowledge Engineering Review, 9(4), 327-354.

Weimer, D., Scholz-Reiter, B., & Shpitalni, M. (2016). Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. CIRP Annals, 65(1), 417-420.

Wen, W., & Xia, A. (1999). Verifying edges for visual inspection purposes. Pattern recognition letters, 20(3), 315-328.

Xie, X. (2008). A review of recent advances in surface defect detection using texture analysis techniques. ELCVIA: electronic letters on computer vision and image analysis, 1-22.

Ye, Y. (2009, August). Fabric defect detection using fuzzy inductive reasoning based on image histogram statistic variables. In 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery (Vol. 6, pp. 191-194). IEEE.

Zbigniew, S. (2004). An introduction to rough set theory and its applications-A tutorial. Proceedings of ICENCO’2004.

Zhang, C., Shi, W., Li, X., Zhang, H., & Liu, H. (2018). Improved bare PCB defect detection approach based on deep feature learning. The Journal of Engineering, 2018(16), 1415-1420.

Zhang, F., & Luk, T. (2007). A data mining algorithm for monitoring PCB assembly quality. IEEE Transactions on Electronics Packaging Manufacturing, 30(4), 299-305.

Zhu, D., Pan, R., Gao, W., & Zhang, J. (2015). Yarn-dyed fabric defect detection based on autocorrelation function and GLCM. Autex research journal, 15(3), 226-232.

DECISION SCIENCES INSTITUTE

School Dropouts in Rural India: Key Household Characteristics and Tipping Points from a Survival Analysis Perspective

Nitin Jain
PhD Student, IIM Udaipur, India
nitin.2020phd@iimu.ac.in

ABSTRACT

Even after seventy-four years of India's independence, unequal access to quality education and high dropout rates in rural Indian schools remain fundamental challenges. The extant research presents a significant opportunity to improve the method of analyzing school dropouts as it is a stylized phenomenon driven strongly by socio-economic factors. While many papers discuss this issue, none of them focus on the grade-wise school dropouts in an integrated manner. This paper addresses this methodological gap by using the survival analysis approach. The findings reveal that the 5th through 7th year of schooling and age group 12-15 years are the critical tipping points for arresting future dropouts, and gender and household characteristics impact the school dropout rates significantly. Therefore, the government should pursue policy instruments along these directions to ensure everyone gets to attend full fourteen years of schooling without disruption.

KEYWORDS: Educational Research, School Dropouts, Survival Analysis, Educational Policy

INTRODUCTION

"*The destiny of India is now being shaped in her classrooms.*"
- India Education Commission, 1964–66

Education is a crucial cog in the wheel of a nation's growth. According to Dreze and Sen (2013), a solid education contributes to individuals' well-being in two ways. It does *intrinsically* by raising individual capabilities and the welfare of future generations through intergenerational transmissions. Whereas *instrumentally,* it does by elevating income and living standards. The "great divergence" between the developed and under-developed nations is said to be reinforced by rapid strides in schooling in developed countries (Chaudhary et al., 2012). Dercon and Krishnan (2009) contend that measures of psychosocial competencies - a sense of inclusion, self-esteem, self-efficacy, and educational aspirations - are associated with caregiver's education and school participation. These play a significant contribution to future socioeconomic status.

Over the last seven decades, India has taken diverse initiatives to improve school education. Be it the "National Policy for Children," adopted in 1974, which considers children a critical asset that needs adequate services (Apte, 1979). Or, in 2001-02, when India launched one of the biggest global educational initiatives – the Sarva Shiksha Abhiyaan (SSA) (Banerji & Mukherjee, 2008). The government recently announced the National Education Policy (NEP)-2020 to make holistic improvements in the education system. However, despite the government's emphasis on equal access to quality education, India has the highest number of illiterates globally, with every third illiterate in the world being an Indian (Jha, 2007). The glaring issues of limited access to and poor quality of education and high dropout rates are key concerns (Mishra, 1999). Surprisingly, in the past, the issues related to educational power have received greater public attention. For example, decisions about which social strata or institutions would benefit from the proposed changes (Rosenthal, 1974).

This paper analyzes an acute crisis in India's primary education in rural areas related to low learning and subsequent dropout. There being no standardized tests until secondary level board exams at grades X and XII, the inadequate learning challenges remain unseen (Banerji, 2000). Government schools fail to provide the most elementary skills (Banerjee et al., 2010). The syllabus assumes linear learning according to which children have mastered the prior grades' competency and skills before progressing to the next grade (Banerji, 2000). For about a decade, until Jan 2019, government schools had an "all pass" policy for primary grades. Thus, mere attendance in a school would lead to promotion to the next grade.

A glaring gap in the extant research on school dropouts in India is the lack of methodological rigor. While increasing over the years, the extant studies are a bit restricted in their methods that have ranged from qualitative approaches to descriptive and regression-based analysis. Further, most of the studies on dropouts are based on administrative surveys. These may have limited information on individual dropout characteristics, dropouts' ways of life, and conduct (Choudhury, 2006), making it imperative to look beyond the above-mentioned traditional methods while analyzing these surveys. To the best of our knowledge, there is no extant research in India that has applied survival analysis in the context of education-related empirical research.

These reasons make it compelling to examine the school dropouts in rural India, where the landscape of education has its unique features, under the lens of survival analysis. Here household and demographic characteristics such as caste, gender, and religion play a pivotal role in a child's educational trajectory. We aim to analyze the following two research questions to identify the tipping points for dropping out of school. These would help in creating an early warning system and reduce the instances of school dropouts. These research questions are essential as there are undesired pecuniary and non-pecuniary outcomes associated with early school-leaving. For example, school dropout is directly associated with poverty, juvenile crimes, a slowdown in the economy, and imposes an increased cost to society (Cabus & De Witte, 2012).

RQ1: Which schooling years are key to reduce dropout rates in rural India?
RQ2: Which are the critical factors that impact student retention in rural Indian schools?

The paper is organized as follows. The next section on background and literature review covers educational reforms, comparison of private and government schools, rural and urban schooling, and school dropouts. After that, the paper discusses the data, succeeded by the methodological contribution of this paper. The analysis, discussion, and conclusion follow in that order.

BACKGROUND & LITERATURE REVIEW

Educational Reforms and Government Incentives

Over the past seven decades, the government has taken many initiatives for the education system in India. According to the Education for All (EFA) program by India, the universalization of elementary education (UEE) was introduced as a composite program of access to education for children up to 14 years of age through formal or non-formal education programs (Acharya, 1994b). UEE includes the provision of facilities, universal enrollment, and universal retention (Acharya, 1994a; Bordoloi, 2011). However, the facilities' provision may not imply universal enrollment, and the latter may not ensure universal retention. Unfortunately, earlier, more emphasis was given on enrollment than on the quality of education, which led to high dropout rates (Chauhan, 2009). With the enactment of RTE in 2009, India started providing free and compulsory education to children in the age group of six to fourteen years (Kaushal, 2012; Rai, 2014). The importance of primary education is enshrined in the Millennium Development Goals (MDG), which called for universal primary education by 2015 (Jayaraman & Simroth, 2015). Because of this, the government started a wide range of policies targeted at increasing school enrollment. The Indian Government introduced the National Program of Nutritional Support to Primary Education in 1995. The latest initiative is the NEP, which is expected to be fully implemented by 2030.

Private versus Government (Public) Schools

Parents consider factors like their income, access to information on schooling, etc. (Chudgar, 2012) to decide on their child's schooling. Caste identity and the number of siblings also influence parents' decision to send their children to private schools (Bhattacharya et al., 2015). Males might be given more preference for expensive private schools (Chudgar & Creed, 2014). Further, there is a negative perception about teaching and infrastructural support in government schools, although teachers in government schools are paid a higher salary (Gouda et al., 2013; Nanjunda & Ramesh, 2010). Private schools have better performance (Desai et al., 2009; Gouda et al., 2013; Goyal, 2007; G. Kingdon, 1996; Tooley & Dixon, 2006; Wadhwa, 2009) as they do not have an "all pass" policy: children are promoted to the next grade based on their performance in regular evaluation, and their parents seek remedial action outside of the school if they do not perform well (Banerji, 2000).

The degree of the gap between private and government schools in rural areas may vary due to factors like the village within which the private school is situated (Chudgar, 2012). Even in the presence of greater private school supply, traditionally disadvantaged children may not be able to avail themselves of these schools, especially at the lower primary level (Chudgar & Creed, 2014). Mediation of schooling by socioeconomic factors raises important questions about equity (Hill et al., 2011; Mehrotra & Panchamukhi, 2006) and discrimination in access (Azam, 2017; Hill et al., 2011). To arrest this inequality, the RTE co-opts private schools for the delivery of education. The state governments now fund twenty-five percent of seats in private schools for children from disadvantaged backgrounds (Kingdon, 2017). Voucher-based reforms, which provide access to private schools, could, therefore, be a potent policy instrument (Tooley, 2016).

Rural versus Urban Schooling

There is a systematic difference in the education standards in rural and urban areas. Rural India has been hampered with chronic failure to deliver quality education to its young population: the enrollment rate is relatively low, and the dropout rate is very high. The key factors impacting access to education in rural areas include teacher absenteeism, lack of quality teachers, remoteness (Mitra et al., 2008); shortage of textbooks, poor infrastructure (Nanjunda & Ramesh, 2010; Pal, 2010); use of rote learning methods (Arora, 2010); walking distance from school, curriculum, poverty, family-level and village-level literacy and awareness, the activity of village panchayats[i] (Caldwell et al., 1985). According to Morrow (2013), children in rural India face contradictions between individualistic values driven by self-development and the formal qualification and the norms and values embedded in their culture, which expects them to support their family via labor. This expectation to support their family steadily diminishes their aspirations to study.

School Dropouts

Although school enrollment has improved in India, school dropout rates are glaringly high (Choudhury, 2006). The world's largest number of out-of-school children are in India, with the majority of these being girls, and yet, the strategies adopted to address dropout rates in girls have tended to be isolated and fragmented (Kaul, 2015). Dropping-out is a process that starts early in development and continues until a student formally withdraws. The process probably begins even before the child enters school, early home environment, and caregiving being very substantial determinants of dropping out (Choudhury, 2006). Therefore, it is not a self-induced decision; parental bonding and family responsibilities played a major role in determining dropout behavior (Choudhury, 2006).

There are manifold reasons for students dropping out of school in India. A herculean task is to create a balanced picture (a pluralist view) of school participation, which integrates different factors that influence it (Drèze & Kingdon, 2001). These factors include economic (opportunity cost of schooling) (Caldwell et al., 1985; Kaul, 2001); poor home conditions (Filmer & Pritchett, 1998; Filmer & Pritchett, 2001; Singh & Khan, 2016; Singh & Mukherjee, 2017); parental education (Chugh, 2011; Dostie & Jayaraman, 2006; Drèze & Kingdon, 2001); geographical (remoteness of schools) (Siddhu, 2011); social (caste effect, bias against females); cultural (schools as a mere representative of cultures without any values); teacher absenteeism; poor teaching infrastructure (Chatterjee et al., 2018); and religious reasons; farm-related and domestic work and distance from school (Caldwell et al., 1985).

The following five school dropout behavior theories have been discussed in the literature (Battin-Pearson et al., 2000).

1. Academic Mediation - Poor academic achievement is one of the strongest predictors of high school dropout.
2. General Deviance - Deviant behavior, which includes delinquency, drug use, etc., impacts the decision to stay put in school.
3. Deviant Affiliation - Bonding with anti-social peers affects school retention.
4. Poor Family Socialization - Family background is overly critical to continue studies.
5. Structural Strains - Demographic factors such as socioeconomic status, ethnicity, gender, etc., are significant determinants of school truancy/ dropout behavior.

Table 1 summarizes the extant literature on school dropouts in India. Most researchers have restricted their methods to regression or descriptive statistics, which are not the right techniques in this context, as discussed later in the section on methodology.

Table 1: Summary of Extant Literature on School Dropouts in India

S.no.	Reference	Data Source/Scope	Method	Findings
1	Banik and Neogi (2015)	Own survey in North-East India	Logistic Regression	School dropout is impacted by earning compulsions, household chores, access to school, school time, parental unawareness.
2	Bashir et al. (2014)	Own survey on females in a district	Descriptive Statistics	Personal and social factors, poor academic performance, and low interest in studies affect school dropout behavior.
3	Basumatary (2017)	Multiple sources	Ordinary Least Squares (OLS)	School dropout depends on the poverty level, distance from school, quality of teachers, social environment, transportation facilities, etc.
4	Baruah and Goswami (2012)	Own survey in Jorhat district of Assam	Descriptive Statistics	Household work, lack of parental guidance, large family size, poor economic condition, and teacher's punishment lead to school dropouts.
5	Choudhury (2006)	Multiple sources	Logistic Regression	Family duties and parental bonding are critical factors.
6	Das (2010)	Case History Analysis	Case History Analysis	Low education status of parents, single-parent families, low self-esteem, grade-detention, and inadequate school atmosphere are associated with school dropouts.
7	Desai et al. (2017)	Own survey on girls in slums of Surat city	Logistic Regression	Only 50% of girls could complete their primary education, and 14% could complete the full schooling. Financial constraints, religion, parental education, etc. are the key drivers affecting school dropout.
8	Gouda and Sekher (2014)	NFHS-3rd round	Logistic Regression	School dropout is high among those belonging to Muslim, Scheduled Caste (SC), and Scheduled Tribe (ST) families.
9	Jayachandran (2007)	NSS-52nd round	Descriptive Statistics	For the "ever-enrolled" children in the age-group 15-19, lack of interest in education is the most crucial factor.
10	Kandasamy and Pramod (2000)	Own survey	FCM	Poor economic conditions, parents' low education, the medium of instruction, and high fees are the key determinants of dropout behavior.
11	Kumar et al. (2017)	Literature survey	Literature survey	Data Mining techniques can help in predicting propensity to leave school.
12	Mukherjee (2010)	Own survey in rural areas of West Bengal	Qualitative	Low level of income, poor infrastructure, lack of consciousness among parents are the key factors leading to school dropout.
13	Pandita (2015)	Ministry of HRD (2002-2011)	Descriptive Statistics	Girls have a higher dropout compared to boys.
14	Reddy and Sinha (2010)	Multiple sources	Descriptive Statistics	Poverty, child labor, household characteristics, and school quality are the significant factors leading to school dropout.
15	Sajjad et al. (2012)	Own survey in south-east Delhi	Descriptive Statistics	Family attributes, income, and education of parents influence school dropout behavior.
16	Sharma et al. (2007)	Own survey in Kangra district of Himachal Pradesh	Descriptive Statistics	A significant association exists between family characteristics, income, mother's education, teachers' behavior, teaching methods, and

				curriculum suitability with the incidence of dropouts.
17	Siddhu (2011)	Own survey in rural areas of Uttar Pradesh	Logistic Regression	The transition from middle to secondary education depends on cost, gender, proximity to school, and being a part of marginalized social groups.
18	Siddiqui (2013)	NSS-43rd round	Descriptive Statistics	Factors impacting school dropout in Muslims include improper attitude towards education, family background, financial constraints, school-related attributes, etc.
19	Sikdar and Mukherjee (2012)	NSS-64th round	Descriptive Statistics	The household atmosphere, financial constraints, and quality of education are the critical factors inducing school dropout.
20	Tabassum (2019)	A survey conducted by TNS	Descriptive Statistics	Boys have a higher propensity to dropout; key factors that lead to dropout include household chores, economic reasons, and migration of families.
21	Yadav et al. (2010)	Own survey in Chandrapur district of Maharashtra	Logistic Regression	Lack of awareness among parents, difficulty in studying due to medium (language) of instruction, teachers' strictness, and failure are the key factors leading to dropout.

Source: Authors' Literature Review

DATA

This paper empirically analyzes the India Human Development Survey (IHDS) - 2011-12. The dataset spans 42,152 households, 204,569 individuals, 4,267 schools from 1,501 villages, and 34 states and union territories in India (Desai & Vanneman, 2015). IHDS is a well-known survey in the extant literature (see, for example, (Azam, 2015; Borooah, 2012; Chatterjee et al., 2018; Chudgar & Creed, 2016; Desai et al., 2015)). It compares well with other national-level surveys and captures a wide range of demographic, socioeconomic, and religious characteristics.

There are different categories of schools in India. For the current analysis, the government schools constitute ESG, government, and government-aided schools. A comprehensive set of control variables have been considered based on the extant literature. These include age, gender, grade, medium of instruction, years of completed education, ever repeated a grade, state, caste category, religion, household per capita consumption, and household highest education. Control variables like the medium of instruction and household per capita income also capture to an extent the impact of school infrastructure. Age is considered as a discrete variable to help a more detailed analysis. This consideration of age as a discrete variable is also one of the methodological improvements over the extant research, which mostly considers age as a continuous variable.

We have tried our best to consider all possible covariates following the five theories on the school dropout mentioned earlier. We also wanted to include control variables such as time spent on household chores but could not consider them due to too many missing values. The final dataset has 29,681 observations, and probability weights (pweights) were used for the analysis. Tables 2-5 show the frequency distribution for the key variables. Females have a higher propensity to dropout. Those from government schools are more inclined to discontinue their studies. Likewise, students belonging to socioeconomically less-privileged backgrounds are more likely to leave school before completing grade XII.

Table 2: Frequency Distribution for Gender versus Dropped Out

Gender (RO3)	Dropped Out (Yes)	Dropped Out (No)
Male	124 (45.09%)	15,582 (52.99%)
Female	151 (54.91%)	13,824 (47.01%)
Total	275	29,406

Note: Column percentages are indicated in parenthesis.

Table 3: Frequency Distribution for School Type versus Dropped Out

School Type	Dropped Out (Yes)	Dropped Out (No)
Government	227 (82.55%)	21,298 (72.53%)
Others	48 (17.45%)	8,108 (27.57%)
Total	275	29,406

Note: Column percentages are indicated in parenthesis.

Table 4: Frequency Distribution for Caste Category versus Dropped Out

Caste Category (ID13)	Dropped Out (Yes)	Dropped Out (No)
Brahmin	9 (3.27%)	1,299 (4.42%)
Forward/General (excl. Brahmin)	34 (12.36%)	5,687 (19.34%)
Other Backward Castes (OBC)	132 (48.00%)	12,213 (41.53%)
Scheduled Castes (SC)	57 (20.73%)	6,879 (23.39%)
Scheduled Tribes (ST)	43 (15.64%)	3,097 (10.53%)
Others	0 (0.00%)	231 (0.79%)
Total	275	29,406

Note: Column percentages are indicated in parenthesis.

Table 5: Frequency Distribution for Religion versus Dropped Out

Religion (ID11)	Dropped Out (Yes)	Dropped Out (No)
Hindu	230 (83.64%)	24,374 (82.89%)
Muslim	33 (12.00%)	3,340 (11.36%)
Christian	4 (1.45%)	517 (1.76%)
Sikh	6 (2.18%)	836 (2.84%)
Buddhist	1 (0.36%)	130 (0.44%)
Jain	0 (0.00%)	16 (0.05%)
Tribal	1 (0.36%)	150 (0.51%)
Others	0 (0.00%)	35 (0.12%)
None	0 (0.00%)	8 (0.03%)
Total	275	29,406

Note: Column percentages are indicated in parenthesis.

METHODOLOGY

An analysis – be it quantitative or qualitative – can only be as good as the underlying data and method. Concerns raised in quantitative studies are around the quality of data, their accuracy (Kingdon, 2008; Kurien, 1981), comprehensiveness, and availability in the desired form at the right time (Aggarwal, 2000). Lack of adequate data is another key limitation cited by most papers (see, for example, Agrawal and Agrawal (2018); Desai et al. (2009)). G. G. Kingdon (1996) argues that enrollments in government-funded elementary schools are greatly exaggerated in official statistics. Therefore, such data issues need to be kept in mind while undertaking analysis on them.

There are issues with the methods also. The most widely used technique, ordinary least squares (OLS), although very simple to apply, has challenges associated with it. A key problem with OLS models is the assumption of normally distributed residuals. Thus, time conditional on x_j is assumed to follow a normal distribution given below (Cleves et al., 2016).

$$time_j \sim N(\beta_0 + \beta_1 x_j, \sigma^2), j = 1,2,3 \dots n \tag{5.1}$$

This is an unreasonable assumption for many events. For instance, if the risk of occurrence of an event is constant over time, wherein the distribution of time would follow an exponential distribution and as such time to failure (dropout from school in this case) is always positive while normal distribution is supported on the full real line. Linear regression is not robust to these violations, so OLS is not appropriate for such analysis. Therefore, the need is to move from the traditional OLS approach to technically more proper methods such as survival analysis, which is briefly described here (for more details, refer to a textbook such as Cleves et al. (2016)).

Let T be a nonnegative random variable denoting the time to failure event (here, it would be dropping out of school). Then, the survivor function is the reverse of the cumulative distribution function of T as shown in (4.2) and reports the probability of survival beyond time t, i.e., the probability there is no failure event (dropping out of school) before t.

$$S(t) = 1 - F(t) = Pr\ (T > t) \tag{5.2}$$

The hazard rate (function), $\mathrm{h(t)}$, is the instantaneous rate of failure and computed as below. It assumes that the individuals at risk and still under observation at failure time are representative of the population risk. This is a fair assumption for our case as the data is from IHDS, a national representative survey, and to a large extent, we could condition on the factors which drive the riskier students to dropout first. Mathematically, we represent the hazard function as:

$$h(t) = \frac{f(t)}{S(t)} \tag{5.3}$$

There is a one-to-one relationship between the probability of survival past a certain time and the amount of risk accumulated until then. The hazard rate measures this risk. The cumulative hazard function, $\mathrm{H(t)}$, measures the amount of risk that has been accumulated until that time, and the following relationships can be drawn.

$$S(t) = \exp\ (-H(t)) \tag{5.4}$$

$$F(t) = 1 - \exp\ (-H(t)) \tag{5.5}$$

$$f(t) = h(t)\exp\ (-H(t)) \tag{5.6}$$

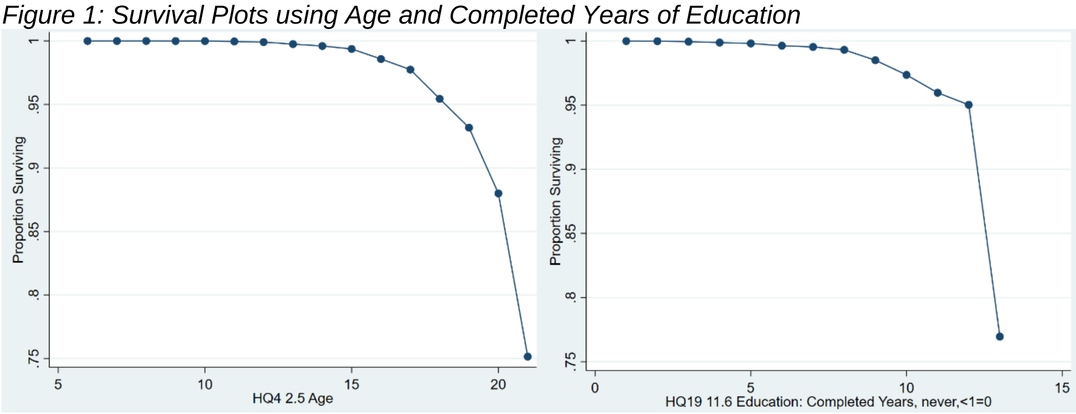
For the current analysis, we first perform a life table analysis (non-parametric) using years of schooling and age as the variables. Then we estimate a parametric model as in (4.1) using Weibull distribution as well as the Cox proportional hazards model, which is given in (4.7).

$$h(t|x_j) = h_0(t)exp\,(x_j\beta_x) \tag{5.7}$$

RESULTS and DISCUSSION

Figure 1 shows the survival plots using Age (in years) and Years of completed education (ED6). The corresponding lifetables for these are shown in Tables 6 and 7.

Figure 1: Survival Plots using Age and Completed Years of Education



Source: Authors' computations

Table 6: Life Tables for Age (in years)

Intervals for Age	Starting Count	Dropouts	Lost	Survival
5-6	29,681	1	1,429	100.00%
6-7	28,251	-	2,038	100.00%
7-8	26,213	-	2,659	100.00%
8-9	23,554	1	2,294	99.99%
9-10	21,259	-	2,168	99.99%
10-11	19,091	6	2,744	99.96%
11-12	16,341	8	2,012	99.91%
12-13	14,321	20	3,027	99.75%
13-14	11,274	15	2,381	99.60%
14-15	8,878	18	2,435	99.37%
15-16	6,425	44	1,779	98.58%
16-17	4,602	32	1,676	97.74%
17-18	2,894	53	1,280	95.44%
18-19	1,561	28	771	93.17%
19-20	762	29	479	88.00%
20-21	254	20	234	75.15%

Source: Authors' computations

Table 7: Life Tables for Completed Years of Education

Intervals for Completed Years of Education	Starting Count	Dropouts	Lost	Survival
0-1	29,681	1	3161	100.00%
1-2	26,519	2	3011	99.99%
2-3	23,506	8	2831	99.95%
3-4	20,667	14	2580	99.88%
4-5	18,073	11	2595	99.81%
5-6	15,467	26	2506	99.63%
6-7	12,935	11	2407	99.54%
7-8	10,517	21	2371	99.31%
8-9	8,125	57	2296	98.50%
9-10	5,772	51	2741	97.36%
10-11	2,980	34	1201	95.97%
11-12	1,745	10	1459	95.03%
12-13	276	29	247	76.96%

Source: Authors' computations

The survival plots and life tables show that age 12-15 is the age-group wherein proactive measures should be taken to prevent school dropout at a later stage. In terms of completed years of education, 5-7 years of education (grades 5-7) is the time when the trigger for dropping out begins. Given the fact that the age in grade 1 is around 6-7 years, both these findings confirm each other well. It is important to note that in both these analyses, we have not used any covariates. Therefore, the next step is performing a parametric analysis, which would help find the household-related factors that affect the completed years of education. We have explored Cox regression as well as Weibull-Proportional Hazards (PH) models the results of which are presented in Table 8.

Table 8: Parametric Analysis Estimates

Variables	Cox Regression (Model 1)	Cox Regression (Model 2)	Weibull-PH regression (Model 3)	Weibull-PH regression (Model 4)
Constant			-6.0877*** (1.3445)	-6.0877*** (1.3445)
CS4_govt – (Whether studying in a government school)				
Yes	0.4456* (0.2369)	0.4306* (0.2438)	0.4537** (0.2204)	0.4307* (0.2340)
ED7 – Ever-repeated a class – Base is No				
Yes	0.7283*** (0.1476)	0.7596*** (0.1444)	0.6724*** (0.1590)	0.7125*** (0.1650)
RO3 (Gender) – Base is Male				
Female	0.2799** (0.1218)	0.2781** (0.1181)	0.2787** (0.123)	0.2766** (0.1203)
RO5 (Age) – Base is 5 years				
6 years		-55.4353 (.)		-29.9104*** (0.9263)
7 years		-57.1019 (.)		-29.9855*** (1.0205)
8 years		-5.7254 (1.1936)		-6.8835*** (1.4082)
9 years		-58.8834 (.)		-29.881*** (0.9843)
10 years		-4.4017*** (0.8005)		-6.2621*** (1.2714)
11 years		-3.9189*** (0.7138)		-6.0556*** (1.2083)
12 years		-4.2585*** (0.6772)		-6.6531*** (1.1450)
13 years		-4.8743*** (0.5475)		-7.4227*** (1.0495)
14 years		-5.266*** (0.6454)		-7.8454*** (1.1137)
15 years		-4.3664*** (0.6101)		-6.9691*** (1.0752)
16 years		-4.7873*** (0.6263)		-7.4151*** (1.0695)
17 years		-4.7343*** (0.6309)		-7.3209*** (1.0918)
18 years		-5.1288*** (0.846)		-7.6616*** (1.2423)
19 years		-4.5936*** (0.7975)		-7.0462*** (1.2385)
20 years		-4.7338*** (0.811)		-7.0858*** (1.2702)
HHEDUC – Base is Never attended				

1st grade	0.1867	0.237	0.2502	0.3155
	(0.5109)	(0.5441)	(0.5127)	(0.5503)
2nd grade	-0.6426*	-0.6793*	-0.6577*	-0.6854*
	(0.3496)	(0.3564)	(0.3714)	(0.3753)
3rd grade	0.0474	0.0551	0.0319	0.0534
	(0.3403)	(0.366)	(0.2901)	(0.3275)
4th grade	-0.7467**	-0.6904*	-0.8058**	-0.7541**
	(0.3585)	(0.3716)	(0.3669)	(0.3827)
5th grade	-0.5476*	-0.5234*	-0.5706**	-0.5377**
	(0.2731)	(0.2767)	(0.2584)	(0.2733)
6th grade	-0.3612	-0.3365	-0.4623	-0.4436
	(0.3777)	(0.3989)	(0.3837)	(0.4143)
7th grade	-1.608***	-1.6485***	-1.5524***	-1.5939***
	(0.2712)	(0.2755)	(0.2884)	(0.2890)
8th grade	-0.7387***	-0.6987***	-0.7902***	-0.7773***
	(0.1794)	(0.1753)	(0.2139)	(0.2088)
9th grade	-1.5487***	-1.5348***	-1.5431***	-1.5438***
	(0.388)	(0.3885)	(0.3901)	(0.3863)
10th grade	-1.375***	-1.354***	-1.4425***	-1.4311***
	(0.3838)	(0.3873)	(0.3797)	(0.3860)
11th grade	-1.3522**	-1.3495***	-1.399**	-1.414***
	(0.5026)	(0.4746)	(0.5012)	(0.4849)
12th grade	-1.1937**	-1.1722***	-1.1842**	-1.1832**
	(0.4305)	(0.4152)	(0.4565)	(0.4535)
Post 12th – Year 1	-1.2504**	-1.2262**	-1.441**	-1.4299**
	(0.5233)	(0.5366)	(0.5717)	(0.5725)
Post 12th – Year 2	-3.0532**	-3.0515**	-3.0432**	-3.0465**
	(1.1144)	(1.1008)	(1.1147)	(1.1086)
Bachelor's	-1.8814***	-1.8754***	-1.8927***	-1.8946***
	(0.5903)	(0.5947)	(0.5750)	(0.5809)
Above Bachelor's	-2.1523***	-2.1274***	-2.2091***	-2.2078***
	(0.7228)	(0.7166)	(0.6816)	(0.6796)
ID11 (Religion) – Base is Hindu				
Muslim	0.5245	0.5292	0.5248	0.536
	(0.3568)	(0.3543)	(0.3601)	(0.3644)
Christian	-0.1523	-0.2473	-0.2364	-0.3485
	(0.8153)	(0.878)	(0.7159)	(0.8177)
Sikh	0.4302	0.396	0.8101**	0.7157**
	(0.3176)	(0.3022)	(0.3057)	(0.2689)
Buddhist	-0.1985	-0.1657	-0.3562	-0.3537
	(0.307)	(0.3183)	(0.3416)	(0.3603)
Jain	-50.1062	-50.3248	-22.2368***	-21.9361***
	(.)	(.)	(0.6611)	(0.5835)
Tribal	-1.1371	-1.0242	-1.2354	-1.1144
	(0.9374)	(0.899)	(0.9704)	(0.9126)
Others	-52.763	-52.6442	-23.9456***	-23.7177***
	(.)	(.)	(1.0857)	(1.0678)
None	-50.6522	-50.6651	-22.7386***	-22.7855***
	(.)	(.)	(0.8888)	(0.9787)
ID13 (Caste) – Base is Brahmin				
General (Except Brahmin)	0.156	0.1784	0.2524	0.2731
	(0.4217)	(0.4345)	(0.4229)	(0.4379)
OBC	0.2803	0.2983	0.2917	0.3024

	(0.4796)	(0.4821)	(0.4879)	(0.4960)
SC	0.2975 (0.5876)	0.3145 (0.5894)	0.3452 (0.5969)	0.3575 (0.6070)
ST	0.2031 (0.565)	0.2161 (0.5791)	0.2196 (0.5565)	0.2104 (0.5726)
Others	-53.0556 (.)	-53.0927 (.)	-22.1791*** (0.5844)	-22.077*** (0.5659)
ln_p			1.4293*** (0.0792)	1.5214*** (0.0746)

Source: Authors' computations, only key variables have been represented due to space constraints
***Significant at 1% level **Significant at 5% level *Significant at 10% level

All the four parametric models indicate that the key attributes of those who have a higher likelihood to dropout are: girls, those having a lower household education and lower per capita consumption, those who have repeated a grade before, and those studying in government schools. Surprisingly, the caste category and religion do not seem to be significant factors here. This insignificance could be because we have also considered state (not shown in the table) and few other variables on household characteristics such as household education, consumption, etc., which may have masked the effect of caste category and religion.

DISCUSSION

Intergenerational persistence is a key aspect of school dropout. Parental education has a significant impact on the education of their children. For India, this is very important due to multiple reasons. The anecdotal notion of family background influencing destiny is very prevalent. There is also a preference for sons over daughters (Kingdon, 2005) and pro-male bias in educational investment. Households that depend critically on their children's incomes for survival lack the ability to respond to the economic benefits of education (Chamarbagwala, 2008). Also, income inequality may lead to the passing of a greater advantage or disadvantage from parents to their children (Azam, 2015).

The impact of primary schooling on economic development is much higher than perceived (Colclough, 1982) but the progress on increasing access to elementary education has been slow, although steady in India (Little, 2010). Therefore, the current analysis holds key importance for school education in rural India. This analysis shows that girls and those whose parents are less educated are at a higher risk of discontinuing their schooling. Further, those in the age group of 12-15 years, as well as 5-7 completed years of schooling education, need to be attended to with proper care and attention. They have a higher propensity to be future dropouts. So, proactive measures need to be adopted to prevent them from leaving their studies. Policy measures need to focus on these aspects to ensure that the goal of a zero percent dropout is met successfully.

To the best of our knowledge, this is the pioneering paper on Indian school dropouts which adopts a survival analysis approach. The next steps could be to consider panel data, explore machine learning approaches, and Bayesian statistics to bring out more informative findings. A shortcoming in the data used here is missing values with some variables such as the time spent on household chores that would be extremely helpful to relate these with the propensity to drop out of school. Future analysis could also help with a wider range of village-related variables to get a more comprehensive picture.

CONCLUSION

School education is a quintessential element for human beings to evolve and contribute to society. Continuity in education gets translated into improved learning outcomes for the students. School education is the foundation to attain higher education. Therefore, it needs to be strong and thorough. There is a growing awareness in academia and practice on raising the quality of schooling. Primary schooling impacts the future trajectory of an individual and society in terms of educational, intellectual, and economic attainment. This paper analyzes an important aspect of the education system in rural India by using survival analysis, which is more appropriate in this context than traditional regression-based methods. It is expected to help academicians, practitioners, and policymakers involved in school education.

REFERENCES

Acharya, P. (1994a). Problems of Universal Elementary Education. *Economic and Political Weekly, 29*(49), 3098-3105.

Acharya, P. (1994b). Universal Elementary Education: Receding Goal. *Economic and Political Weekly, 29*(1/2), 27-30.

Aggarwal, Y. (2000). *Public and Private Partnership in Primary Education in India*. National Institute of Educational Planning and Administration.

Agrawal, T., & Agrawal, A. (2018). Who Gains More from Education? A Comparative Analysis of Business, Farm and Wage Workers in India. *The Journal of Development Studies, 55*(6), 1081-1098. https://doi.org/10.1080/00220388.2018.1443209

Apte, M. (1979). Priority Needs of Children in India. *Indian Journal of Public Administration, 25*(3), 549-568.

Arora, P. (2010). Hope-in-the-Wall? A digital promise for free learning. *British Journal of Educational Technology, 41*(5), 689-702. https://doi.org/10.1111/j.1467-8535.2010.01078.x

Azam, M. (2015). Intergenerational Educational Persistence among Daughters: Evidence from India. *IZA Discussion Papers*.

Azam, M. (2017). Explaining Caste Differences in Private School Attendance. *Review of Development Economics, 21*(4), 1191-1204. https://doi.org/10.1111/rode.12305

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy, 2*(1), 1-30.

Banerji, R. (2000). Poverty and primary schooling: field studies from Mumbai and Delhi. *Economic and political weekly*, 795-802.

Banerji, R., & Mukherjee, A. N. (2008). Achieving Universal Elementary Education in India. *Margin: The Journal of Applied Economic Research, 2*(2), 213-228. https://doi.org/10.1177/097380100800200204

Banik, A., & Neogi, D. (2015). Poverty and earning compulsions for the family have pushed children out of schools. *International Journal of Social Economics*.

Baruah, S. R., & Goswami, U. (2012). Factors influencing school dropouts at the primary level. *International Journal of Farm Sciences, 2*(2), 141-144.

Bashir, A., Khan, N., & Vaida, N. (2014). Personal, School and Family-main factors affecting female school dropout. *Journal of Scientific and Innovative Research, 3*(3), 295-298.

Basumatary, M. (2017). *A study on causes of school dropouts in baganpara circle of baksa district, Assam* Assam Don Bosco University].

Battin-Pearson, S., Newcomb, M. D., Abbott, R. D., Hill, K. G., Catalano, R. F., & Hawkins, J. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of educational psychology, 92*(3), 568.

Bhattacharya, S., Dasgupta, A., Mandal, K., & Mukherjee, A. (2015). Understanding the "Sorting Hat": The Role of Family and Caste Network in School Choice Decision. *ESP Working Paper Series*(69).

Bordoloi, R. (2011). Challenges in Elementary Education in India: Various Approaches. *Journal of Education and Practice, 2*(7).

Borooah, V. K. (2012). Social Identity and Educational Attainment: The Role of Caste and Religion in Explaining Differences between Children in India. *Journal of Development Studies, 48*(7), 887-903. https://doi.org/10.1080/00220388.2011.621945

Cabus, S. J., & De Witte, K. (2012). *School Absenteeism and Dropout, A Bayesian Duration Model*.

Caldwell, J. C., Reddy, P., & Caldwell, P. (1985). Educational transition in rural South India. *Population and Development Review*, 29-51.

Chamarbagwala, R. (2008). Regional Returns to Education, Child Labour and Schooling in India. *The Journal of Development Studies, 44*(2), 233-257. https://doi.org/10.1080/00220380701789935

Chatterjee, I., Li, I., & Robitaille, M.-C. (2018). An overview of India's primary school education policies and outcomes 2005–2011. *World Development, 106*, 99-110. https://doi.org/10.1016/j.worlddev.2018.01.016

Chaudhary, L., Musacchio, A., Nafziger, S., & Yan, S. (2012). Big BRICs, weak foundations: The beginning of public elementary education in Brazil, Russia, India, and China. *Explorations in Economic History, 49*(2), 221-240. https://doi.org/10.1016/j.eeh.2012.02.001

Chauhan, C. P. S. (2009, 2009/03/01). Education for all in India: A second look. *International Journal of Lifelong Education, 28*(2), 227-240. https://doi.org/10.1080/02601370902757091

Choudhury, A. (2006). Revisiting Dropouts: Old Issues, Fresh Perspectives. *Economic and Political Weekly, 41*, 5257-5263.

Chudgar, A. (2012). Variation in Private School Performance: The Importance of Village Context. *Economic and Political Weekly, 47*(11), 52-59.

Chudgar, A., & Creed, B. (2014). *How does Demand for Private Schooling Vary Across Locations with Different Private School Supply? Analysis of Data from Rural India*. Michigan State University.

Chudgar, A., & Creed, B. (2016). How are private school enrolment patterns changing across Indian districts with a growth in private school availability? *Oxford Review of Education, 42*(5), 543-560. https://doi.org/10.1080/03054985.2016.1217688

Chugh, S. (2011). *Dropout in secondary education : a study of children living in slums of Delhi*.

Colclough, C. (1982). The Impact of Primary Schooling on Economic Development: A Review of the Evidence. *World Development, 10*(3), 167-185.

Das, P. (2010). Process of girls dropout in school education: Analysis of selected cases in India. Engendering Empowerment: Education and Equality e-Conference, April,

Dercon, S., & Krishnan, P. (2009). Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries. *Children, Youth and Environments, 19*(2), 138-163.

Desai, K., Moitra, M., Chaudhary, V., & Dave, B. (2017). Education status, school dropouts and its reasons, determinants and perspectives among young girls of a city of Western India. *International Journal of Medical Science and Public Health, 6*(8), 1261-1267.

Desai, S., Dubey, A., Banerji, R., & Vanneman, R. (2009). Private schooling in India: A new educational landscape. In S. Bery, B. Bosworth, & A. Panagariya (Eds.), *India Policy Forum 2008-09* (Vol. 5, pp. 1). SAGE Publications India.

Desai, S., Joshi, O., & Vanneman, R. (2015). Do Public Works Programs improve Child Education: Evidence from India. *ASER*.

Desai, S., & Vanneman, R. (2015). India human development survey-ii (ihds-ii), 2011-12. *ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 31*.

Dostie, B., & Jayaraman, R. (2006). Determinants of School Enrollment in Indian Villages. *Economic development and cultural change, 54*(2), 405.

Drèze, J., & Kingdon, G. G. (2001). School Participation in Rural India. *Review of Development Economics, 5*(1), 1-24.

Dreze, J., & Sen, A. (2013). *India: Development and Participation* (2 ed.). Oxford University Press.

Filmer, D., & Pritchett, L. (1998). *Educational enrollment and attainment in India: Household wealth, gender, village, and state effects*. World Bank Washington, DC.

Filmer, D., & Pritchett, L. H. (2001, February 01). Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application To Educational Enrollments In States Of India [journal article]. *Demography, 38*(1), 115-132. https://doi.org/10.1353/dem.2001.0003

Gouda, J., Chandra Das, K., Goli, S., & Maikho Apollo Pou, L. (2013). Government versus private primary schools in India. *International Journal of Sociology and Social Policy, 33*(11/12), 708-724. https://doi.org/10.1108/ijssp-12-2012-0105

Gouda, S., & Sekher, T. (2014). Factors leading to school dropouts in India: An analysis of national family health survey-3 data. *IOSR Journal of Research & Method in Education, 4*(6), 75-83.

Goyal, S. (2007). Learning achievements in India: A study of primary education in Rajasthan. *Human Development Unit, South Asia Region, The World Bank, Washington DC*.

Hill, E., Samson, M., & Dasgupta, S. (2011). Expanding the school market in India: Parental choice and the reproduction of social inequality. *Economic and political weekly*, 98-105.

Jayachandran, U. (2007). How High Are Dropout Rates in India? *Economic and Political Weekly, 42*(11), 982.

Jayaraman, R., & Simroth, D. (2015). The Impact of School Lunches on Primary School Enrollment: Evidence from India's Midday Meal Scheme Impact of school lunches on primary school enrollment. *The Scandinavian Journal of Economics, 117*(4), 1176-1203.

Jha, P. (2007). Guaranteeing Elementary Education: A Note on Policy and Provisioning in Contemporary India. *Journal of South Asian Development, 2*(1), 75-105. https://doi.org/10.1177/097317410600200104

Kandasamy, W. V., & Pramod, P. (2000). Parent Children Model using FCM to Study Dropouts in Primary Education. *Ultra Sci, 13*, 174-183.

Kaul, R. (2001). Accessing primary education: Going beyond the classroom. *Economic and Political Weekly*, 155-162.

Kaul, R. (2015). Gender Inequality: Challenges of Educating the Girl Child. *Social Change, 45*(2), 224-233.

Kaushal, M. (2012). Implementation of Right to Education in India: Issues and Concerns. *Journal of Management and Public Policy, 4*(1), 42-48.

Kingdon, G. (1996). The Quality and Efficiency of Private and Public Education: A Case Study of Urban India. *Oxford Bulletin of Economics and Statistics, 58*(1).

Kingdon, G. (2008). School-sector Effects on Student Achievement in India. In R. Chakrabarti & P. E. Peterson (Eds.), *School Choice International: Exploring Public-Private Partnerships* (pp. 111-139). MIT Press.

Kingdon, G. G. (1996). Private Schooling in India: Size, Nature, and Equity-Effects. *Economic and Political Weekly, 31*(51), 3306-3314.

Kingdon, G. G. (2005). Where has all the bias gone? Detecting gender bias in the intrahousehold allocation of educational expenditure. *Economic Development and Cultural Change, 53*(2), 409-451.

Kingdon, G. G. (2017). *The Private Schooling Phenomenon in India: A Review* (10612). Institute of Labor Economics (IZA).

Kumar, M., Singh, A., & Handa, D. (2017). Literature survey on educational dropout prediction. *International Journal of Education and Management Engineering, 7*(2), 8.

Kurien, J. (1981). Towards Universal Elementary Education: Promise and Performance. *Economic and Political Weekly, 16*(40), 1608-1618.

Little, A. W. (2010). *Access to Elementary Education in India: Politics, Policies and Progress* (CREATE, Issue.

Mehrotra, S., & Panchamukhi, P. R. (2006). Private provision of elementary education in India: findings of a survey in eight states. *Compare: A Journal of Comparative and International Education, 36*(4), 421-442. https://doi.org/10.1080/03057920601024883

Mishra, R. (1999). Research on education in India. *Prospects, 29*(3), 335-347.

Mitra, S., Dangwal, R., & Thadani, L. (2008). Effects of remoteness on the quality of education: A case study from North Indian schools. *Australasian Journal of Educational Technology, 24*(2), 168-180.

Morrow, V. (2013). Whose Values? Young People's Aspirations and Experiences of Schooling in Andhra Pradesh, India. *Children & Society, 27*(4), 258-269. https://doi.org/10.1111/chso.12036

Mukherjee, D. (2010). Reducing out of school children in India: lessons from a micro study.

Nanjunda, D. C., & Ramesh, C. (2010). Elementary Education in Rural Karnataka - An Analysis. *Artha Journal of Social Sciences, 9*(2), 1-11.

Pal, S. (2010). Public infrastructure, location of private schools and primary school attainment in an emerging economy. *Economics of Education Review, 29*(5), 783-794. https://doi.org/10.1016/j.econedurev.2010.02.002

Pandita, R. (2015). Enrolment & dropout percentage among boys & girls up to secondary level in India: A comparative study. *International Letters of Social and Humanistic Sciences, 49*, 123-134.

Rai, A. (2014). Implementation of the RTE Act: RTE Forum's Stocktaking Report. *Social Change, 44*(3), 439-449. https://doi.org/10.1177/0049085714536806

Reddy, A. N., & Sinha, S. (2010). *School Dropouts or Pushouts? Overcoming Barriers for the Right to Education*.

Rosenthal, D. B. (1974). Educational Politics and Public Policymaking in Maharashtra, India. *Comparative Education Review, 18*(1), 79-95.

Sajjad, H., Iqbal, M., Siddiqui, M. A., & Siddiqui, L. (2012). Socio-economic determinants of primary school dropout: evidence from south east Delhi, India. *European Journal of Social Sciences, 30*(3), 391-399.

Sharma, R., Sharma, S., & Nagar, S. (2007). Extent of female school drop outs in Kangra District of Himachal Pradesh. *Journal of Social Sciences, 15*(3), 201-204.

Siddhu, G. (2011). Who makes it to secondary school? Determinants of transition to secondary schools in rural India. *International Journal of Educational Development, 31*(4), 394-401. https://doi.org/10.1016/j.ijedudev.2011.01.008

Siddiqui, M. H. (2013). The Problems of School Drop Outs among Minorities with Special Reference to Muslims in India. *International Journal of Management and Social Sciences Research (IJMSSR) ISSN: 2319-4421 Volume, 2*.

Sikdar, S., & Mukherjee, A. N. (2012). Enrolment and dropout rate in school education. *Economic and Political Weekly*, 27-31.

Singh, R., & Khan, S. (2016). Perspectives on Children's Work and Schooling: Evidence from a Longitudinal Study in Andhra Pradesh and Telangana, India. *ILO Asia-Pacific Working Paper Series*.

Singh, R., & Mukherjee, P. (2017). Diverging Pathways: When and Why Children Discontinue Education in India. *Young Lives*.

Tabassum, S. R. (2019). A Study of School Dropouts in India: Sensitivity at Display. *Journal of Applied Linguistics and Language Research, 6*(3), 1-10.

Tooley, J. (2016, 2016/09/02). Extending access to low-cost private schools through vouchers: an alternative interpretation of a two-stage 'School Choice' experiment in India. *Oxford Review of Education, 42*(5), 579-593. https://doi.org/10.1080/03054985.2016.1217689

Tooley, J., & Dixon, P. (2006). 'De facto' privatisation of education and the poor: implications of a study from sub-Saharan Africa and India. *Compare: A Journal of Comparative and International Education, 36*(4), 443-462. https://doi.org/10.1080/03057920601024891

Wadhwa, W. (2009). Are Private Schools Really Performing Better than Goverment Schools? *ASER*.

Yadav, S., Kalakoti, P., & Ahmad, N. (2010). Determinants of school dropouts in children. *Review of Global Medicine and Healthcare Research, 1*(1), 232-245.

[i] Panchayat is the oldest system of local government in the Indian subcontinent and dates back to the 250 CE period. Traditionally, panchayats comprised of wise and respected elders chosen and accepted by the local community. They would settle disputes between individuals and between villages.

DECISION SCIENCES INSTITUTE
e-Government: Focus India

Vivek Shah
Texas State University
email: vs01@txstate.edu

Kamlesh Mehta
National University
kmehta@nu.edu

ABSTRACT

Globally, the move towards e-Government and innovations in e-Governance are occurring rapidly. This paper attempts to understand the role of e-Government and discuss the issues, benefits, and application of the successful implementation of e-Government around the world with a special focus on India. The government of India has launched several projects in support of e-Governance, such as e-seva, smart government, and digital India. Each of these projects are supposed to benefit the citizens; however, the success of these initiatives is not apparent.

KEYWORDS: India, e-Government, e-Governance, Globalization

INTRODUCTION

Globally, governments are faced with the challenges associated with providing efficient and timely services to their citizens. The use of smart technology, artificial intelligence, and blockchain is on the rise. Therefore, the creation of a smart government is essential to reduce interactions through effective and efficient services. The resulting outcomes and trends are a rapid growth toward E-Government and innovations in e-Governance. As a result of smart technologies, the type and volume of the transaction have exponentially grown. Today, the scope of e-Government is enormous, and public services provided through E-Government include numerous interactions and transactions as follows: (1) government-to citizens, (2) government-to-business, (3) government-to-government, and (4) government-to-employees.

LITERATURE REVIEW

The literature review reveals that many research studies have identified several issues concerning e-Governance, which include implementation, smart technologies concerning public sector management, service delivery in relation to no-stop shop, distrust, accountability, cloud computing, and so on.

Criado and Gil-Garcia (2019) seem to advocate for public value generation using smart technologies and strategies in public sector management. The key argument presented in their study is that smart technologies have the potential to foster the co-creation of public services and the generation of public value in management processes, based on the collaborative, social, and horizontal nature of these smart technologies. The study concludes that new smart technologies and strategies will shape and be shaped by the future of public organizations and management.

Scholta et al. (2019) observed two issues facing the current government service delivery: (1) Citizens perceive forms as cumbersome in a single points-of-contact in one-stop shops, and (2) citizens expect government proactively provide appropriate government services themselves, instead of relying on requests for services from citizens. To address these two issues, the authors propose a transition from a one-stop-shop to a no-stop shop, where the citizen does not have to perform any action or fill in any forms to receive government services.

In a study by Hole (2016), a population model shows distrust in e-Government services--the belief that the service deliberately sabotages the user's intent--spreads rapidly even when problems affect only a few users. Strategies such as user-focused development and the use of cloud-based services can help mitigate that distrust.

Tadili and Semma (2015) have explored the use of Cloud Computing in e-Government, identified challenges and benefits of this use, and identified countries that have followed this approach.

Al-Shbail and Aman (2018) provide an understanding on how e-Government enhances public organizations' accountability by highlighting the key elements of e-Government that mitigate the disorders and dysfunctions of accountability relationships.

THE PURPOSE

The purposes of this paper are: (1) to understand the terms e-Government and e-Governance, (2) to understand the role of e-Government in India, (2) challenges to the successful implementation of e-Government, (3) to discuss the issues, benefits, and applications of e-Government, and (4) to explore whether the advent and diffusion of online media, information and communication technologies and their quintessence into eGovernment could be an important factor in India.

E-GOVERNEMNT AND E-GOVERANCE

In the earlier part of the 21st century, the focus of the e-Government was technology. However, over the past decade, the focus of e-Government has shifted from technology to management, including scope, service quality, communication between the citizens and government, and policies and processes.

e-Government definition ranges from "the use of information technology to free movement of information to overcome the physical bounds of traditional paper and physical based systems" to "the use of technology to enhance the access to and delivery of government services to benefit citizens, business partners and employees" (wikibooks, 2021).

In this paper, the authors describe these two terms as follows: e-Government (or electronic government) uses electronic communications devices, such as computers and the Internet, to provide public services to citizens and other persons in a country or region. e-Governance is associated with carrying out the functions and achieving the results of governance through the utilization of information and communication technology (ICT).

FOCUS: INDIA

India is emerging as a global IT superpower. However, it is a paradox that, as a nation, India is far away from the IT revolution. The private sector is leading the drive in the IT revolution, not the government. India has the capabilities of transforming the country into an online nation. What is stopping this from becoming? Despite many successful government IT projects, India has not transformed itself into a digital nation.

India faces a magnitude of challenges.

- Over the last few years, growth and development have been comparatively slow, with per capita income stagnating and not keeping up with the population growth rates.
- A significant 'digital divide' in media, information, and communication technologies has emerged and growing.
- It is vulnerable to energy price fluctuations.
- Environmental concerns are growing, such as pollution.
- India faces increasing water shortages due to global climate change and potential water-rights related conflicts internally and externally.
- Large military expenditures due to border conflicts with neighboring countries.
- A weakness of governance and a dominant role of the public sector in economic activity, despite the beginning of reform and privatization.
- India displays a low level of international and global economic integration.
- India displays a modest level of international economic integration in an era when technology is imposing extensive economic & financial linkages and increased globalization.

Given these challenges and the prominent role of the Indian Government in economic activity, this paper explores whether the advent and diffusion of online media, information, and communication technologies and their quintessence into eGovernment could be an important factor in India.

The acceleration of globalization over the last decade of the 2000s, with increased economic and financial interdependence, substantially impacts policymaking, restricting the scope for autonomous policy decisions. Increased globalization has raised series of new challenges. With the expanded open markets, price transparency due to the Internet, free flow of capital, countries compete for the same resources and goods. Where does India stand on the globalization scale?

Integration with the global economy, measured by the extent of international trade, is relatively high for India, with the ratio of total trade to GDP running at 40.77% and 43.13% for 2017 and 2018, respectively (Microtrends). However, there is a failure to attract foreign financial resources, foreign direct investment, and capital flows.

THE ROLE OF e-GOVERENMENT IN INDIA

India has large public sectors. Many regional political parties run different states with their self-interest and complex regulatory structure and regulations. The development and modernization of public institutions have not been a priority, and limited attention has been given to promote participation and competition. The issue is how and whether e-Government initiatives can help governments achieve better results in realizing their policy goals. Countries around the world, from low income to the highly industrialized, are adopting the Internet. The potential benefits of the network readiness of government are as substantial and varied as the importance of government in the lives of individuals, citizens, and businesses.

It is meant to be a more accessible and contemporary alternative to how people have always done it. The e-Governance should be transparent, faster, and equitable access to all the participants. With increased and evident openness of goods markets and unlimited capital inflow and outflow, countries compete in the global market for the same investments, human skills, and the latest technological innovations.

e-Governance should be applied to all government functions and services. This includes the exchange of information, communications, services, documents and transactions between government and other governments, inter-government as well as citizens. In principle, e-Governance should bring decentralization and democratization. Technology has spread throughout the world over the last decade, and it has driven changes in the industrial and economic model.

Technological changes have made profound social and political impacts which in turn transform decision-making at every level of government. Government is by nature an information intensive organization. A democracy relies on the notion that the best way to make a decision is through the wide participation of all its citizens. ICT facilities two-way communication and relationship between citizens and government agencies. It is imperious that governance should be based upon transactions in the virtual space in the digital economy framework.

In this information-rich and global society, issues regarding decision-making in social, political, and organizational contexts need to be considered. The issue is how and whether e-Government initiatives can help governments achieve better results in realizing their policy goals. The potential benefits of the successful implementation of e-Government are substantial and varied as the importance of government in the lives of individuals and businesses (Saidi & Yared, 2003).

- Elimination of barriers: to overcome the physical and virtual isolation of individuals and communities, allowing better information about the policies and processes of both central and local government.
- Promotion of efficiency: through reduced transactions costs and time allocated, and the streamlining of services and processes.
- Creation of opportunities: citizens, individuals, and businesses have a wider choice, more convenience, and freedom to choose with government processes and services available 24 hours. You can go online, instead of in-line!
- Minimization of waste, elimination of graft, bribery and corruption, and to increase the transparency of government.

- Promote and enhance democracy through increased and improved information and direct participation.

CHALLENGES TO A SUCCESSFUL IMPLEMENTATION OF E-GOEVERENMENT

For any governments to reap the full potential of benefits of e-Government, several conditions and several building blocks are necessary: leadership, connectivity and network readiness, business environment, human capital, privacy, trust, and security.

- Are the leadership and strategic thinking ready?
  e-Government initiatives require substantial investments and innovation before they are prepared for public use. e-Government implies government reform as part of the business process re-engineering of government. Strong, high-level leadership at both central government and state government is needed to supply the vision, establish a national e-Strategy, and support from the public and private sectors.

- Access, Connectivity & Network Readiness – is the technological infrastructure ready?
  Reliable, accessible, and secure telecommunication infrastructure is key to the development of e-Government. This includes the availability of telecommunication services, community access centers, and network readiness. The main challenges are the pricing, affordability, and reliability of network access.

  India has decentralized voice telephony, whereas data and media communications are most important for a knowledge-based economy. The digital divide is the continuum of the Internet and other digital media that separates those who have access to such media from those who do not have access to those media. Approximately 630 million people, half of the India's population of 1.2 billion people have Internet subscriptions. However, other half does not have access to the Internet and most probably live in the rural area. The divide is shaped both by the availability of internet services in different regions and the ability of individuals to tap into those services. A person's location, income, gender, education, language, and age are some of the factors that define their access. One of the India's flagship programs, the Digital India project, identifies universal access to mobile connectivity as one of its main pillars. (Parsheera, 2019).

  Further, the cost of setting up a large-scale technological infrastructure has declined rapidly. For India, there is an advantage from being adapters of following the leader: they can benefit from the experience of what to do, and more importantly, of what not to do. They can and should leapfrog and avoid legacy technologies.

- E-business climate – is the institutional infrastructure ready?
  E-Government can only succeed in an environment where the appropriate framework supports the development of e- initiatives. Political stability, financial soundness, government policies, and enforceability will affect competitiveness in e-business and the network readiness of India. A comprehensive, coordinated approach should include: Economic and financial policy, an open attitude towards investment and trade, a financial system ready to invest into e-business ventures, and promoting e-business itself by integrating it with electronic payment systems, e-payments, Predictable regulation and enforceability, Legal protection and enforcement of intellectual property rights.

- Human capital: Is the population e-aware and e- ready? A major building block, a critical factor for success, is a country's human capital. The population needs to use technology and understand, create local content, and manage e-initiatives.

  Governments through education policies should influence the formation of human capital. India has an educated workforce, but they are not able to capitalize on this.

  India needs to reform educational curricula to include computer literacy. High-quality computer training opportunities should be set up, including on-the-job training and distance learning. Computer education should start in primary schools, thus laying the foundations for a high computer skill level in the population, enhancing the creation of e- human capital.

  India's literacy rate has increased from 69.3% in 2011 to 74.37% in 2018 (Microtrends). However, India is not able to capitalize on an educated workforce or human capital. In 2017, 60% of engineering students remained unemployed or underemployed. 950,000 Indian Scientists and Engineers lived in the US alone in 2013 (Trines, 2018).

- Trust, Information security and privacy: is the legal infrastructure ready? Without providing trust and security for its citizens and businesses, e-Government will not achieve its potential objectives. The confidentiality, integrity, and security of data transmission, the processing, and storage of networked information need to be trusted. This includes strengthening the legal framework to address privacy protection and prosecution of computer crimes, create and enforcing certification authorities.

  The legal infrastructure to support and implement e-Government should include a core set of laws for data protection, electronic signature of communication, cybercrime and computer misuse law, etc. There are many laws in India; however, navigating through the judicial system in India is not that easy.

The government of India has launched several projects in support of e-Governance, such as e-seva, smart government, digital India, e-kranthi, etc. (Nagaraga, 2016). Each of these projects is supposed to benefit the citizens; however, none of these projects have yielded needed results.

e-GOVERNMENT AND e-GOVERNANCE: ISSUSES AND APPLICATIONS

Many studies have researched various elements of e-Government around the world. These studies have identified countries that are engaged in e-Government.

Kurfali et al., (2017) investigated underlying factors that play a role in citizens' decision to use e-Government services in Turkey. According to the results, Performance expectancy, Social influence, Facilitating Conditions, and Trust in the Internet were found to affect behavioral intention to use e-Government services positively.

Bevacqua and Renolds (2019) have researched digitization of taxpayer services as part of the more general trend toward E-Government in the USA and cautioned that ax authorities must be conscious of a significant and rapidly evolving existence 'digital divide' between various demographic groups.

Kumar and Mukherjee (2017) investigated how citizens experience e-Gov services and how that experience influences their behavior in India. The study also investigates how users value e-Gov services in India. The findings highlight the extent to which citizens are moving from traditional ways of using government services to using it electronically. The study also considers the effect of these technological innovations in government settings from a participant's perspective.

Joo and Hovav (2016) analyzed the determinants of information security that influence the adoption of Web-based integrated information systems (IIS) by government agencies in Peru. The study introduces Web-based information systems designed to formulate strategic plans for the Peruvian government. Findings suggest that organizations should implement preventive efforts by introducing various information security solutions and improve information security awareness while reducing perceived information security threats.

In their research study, Zubareva and Byelov (2015) investigated the main concepts of quality of service influencing the operational security of the electronic government system of Ukraine. They analyzed the parameters and safety requirements that ensure the reliable operation of the e-Government system.

One research study ranked eight critical success factors (CSFs) of e-Government as follows: (1) financial situation of government units, (2) public outlay on hardware, networks, and telecommunications, (3) electronic communication between units, (4) integration of front-office and back-office information systems (5) top management support, (6) ICT competences of government employees, (7) state standardization of e-Government solutions, and (8) information security in government (Ziemba, et al., 2016).

In their study, Wirtz and others (2017) have summarized the drivers of e-Government implementation into three categories: (1) improvement of service quality, (2) improvement of internal administration processes, and (3) cost and time savings. The study further identified the barrier to e-Government implementation into four categories as follows: (1) loss of a contract to citizens, (2) IT infrastructure, (3) security risks, and (4) overburdening of administrative staff.

Table 1 lists five most common issues of concern and most common benefits related to e-Government and e-Governance.

Table 1: e-Government and e-Governance – The Common Issues and Benefits

	The Common Issues		*The Common Benefits*
1	Security, Privacy, and Trust	1	Access to information
2	Hardware and software infrastructure	2	Transparency
3	Digital divide amongst citizens	3	Accountability
4	Quality of service	4	Savings in cost, time, and efforts
5	Efficient communication between the citizens and the government	5	Improved performance of government agencies

CONCLUSIONS

Globally, the move towards E-Government and innovations in e-Governance are occurring rapidly. During the past decade, the focus of e-Government has shifted from technology to management including scope, service quality, digital divide, communication between the citizens and government, and policies and processes.

The successful implementation of e-Government and e-Governance is very common in the western hemisphere. As in any technological innovation or project, challenges are inherent. However, once the challenges are conquered, the benefits of e-Government and e-Governance often outweigh the drawbacks.

In spite of many successful government IT projects, the India has not transformed itself into a digital nation. India faces a magnitude of challenges. The government of India has launched several projects in support of e-Governance, such as e-seva, smart government, digital India, and e-kranthi. Each of these projects are supposed to benefit the citizens, however, the success of these initiatives is not clear.

REFRENCES

Al-Shbail, T., & Aman, A. (2018). E-government and accountability: How to mitigate the disorders and dysfunctions of accountability relationships. *Transforming Government*: People, Process and Policy, 12(2), 155–190.

Bevacqua, J., & Renolds, V. (2019). The digital divide and taxpayer rights – cautionary findings from the United States. *eJournal of Tax Research*, 16(3), 714–736.

Criado, J., & Gil-Garcia, J. (2019). Creating public value through smart technologies and strategies: From digital services to artificial intelligence and beyond*. International Journal of Public Sector Management*, 32(5), 438–450. https://doi.org/10.1108/IJPSM-07-2019-0178

Hole, K. (2016). Building Trust in E-Government Services. *Computer*, 49(1), 66–74.

Joo, J., & Hovav, A. (2016). The influence of information security on the adoption of web-based integrated information systems: an e-government study in Peru. Information Technology for Development. *Information Technology Development*, 22(1), 94–116.

Kaufman*n,* D., and Kraay A. (2017). Measuring the Ups and Downs of Governance. https://www.brookings.edu/blog/future-development/2017/09/22/measuring-the-ups-and-downs-of-governance/, September 22.

Kumar, R., Sachan, A., & Mukherjee, A. (2017). Qualitative approach to determine user experience of e-government services. *Computers in Human Behavior*, 71.

Kurfali, M., Arifoğlu, A., Tokdemir, G., & Paçin, Y. (2017). Adoption of e-government services in Turkey. *Computers in Human Behavior*, 66, 168-178.

Mircotrends. India trade to GDP ratio 1996-2020. https://www.macrotrends.net/countries/IND/india/trade-gdp-ratio,”

Parsheera S. (2019). India's on a Digital Sprint that is Leaving Millions Behind. https://www.bbc.com/news/world-asia-india-49085846, October 17, 2019.

Saidi N., and Yared H. (2003). E-Government: Technology for Good Governance, Development and Democracy in MENA Countries. Working Paper 0304, February. https://erf.org.eg/publications/e-government-technology-for-good-governance-development-and-democracy-in-the-mena-countries/

Scholta, H., Mertens, W., Kowalkiewicz, M., & Becker, J. (2019). From one-stop shop to no-stop shop: An e-government stage model. *Government Information Quarterly*, 2019-01, Vol.36 (1), p.11-26.

Tadili, H., & Semma, A. (2015). How governments can benefit from cloud computing. *International Journal of Computer Science Issues*, 12(5), 170-174.

Trines, E. (2018). Education in India. https://wenr.wes.org/2018/09/education-in-india, September 13, 2018.

Wikibooks. (2021). E-government wikibooks URL: https://en.wikibooks.org/wiki/E-government/

Wirtz, B., Weyerer, J., Thomas, M., & Möller, A. (2017). E-Government Implementation: Theoretical Aspects and Empirical Evidence. *Public Organization Review*, 17(1), 101–120.

Ziemba, E., Papaj, T., Zelazny, R., & Jadamus-Hacura, M. (2016). Factors Influencing The Success Of E-Government. The Journal of Computer Information Systems, 56(2), 156–167.

Zubareva, O., & Byelov, S. (2015). Electronic Government System of Ukraine and a Method for increasing Its Security. *Cybernetics and Systems Analysis*, 51(3), 481–488.

DECISION SCIENCES INSTITUTE

Is Localization better than Globalization? Evidence from the Nutraceuticals for Indian Malnourished Children

Shailly Chaurasia, Rupesh Kumar Pati, Sidhartha S. Padhi
Quantitative Methods and Operations Management, Indian Institute of Management Kozhikode, Kerala, 673570, India
Email: shaillyc10fpm@iimk.ac.in

ABSTRACT

The Socio-Economic impacts of the global burden of malnutrition has been observed to be serious with long term impact on individuals and the countries affected by it. Despite of being one of the largest producers of many agricultural products, India's performance in Global Hunger Index has consistently been dismal. World Food Programme (WFP) allocates budget so as to provide a Nutraceutical Product (Plumpy Nut) in developing/under developing nations to reduce the proportion of malnourished children. This study proposes linear programming model towards designing a sustainable nutraceutical supply chain network. The model attempts to investigate if localizing the availability of the nutraceutical products vis-à-vis global sourcing could be looked as a potential solution towards better management of malnutrition, which heavily depending on financial support from various sources. The study helps in the achievement United Nations Sustainable development goals-2030 of *Zero hunger (SDG-2)* and *Responsible Consumption-Production* (SDG-12).

KEYWORDS: Nutraceutical Supply Chain (NSC), Plumpy Nut, Localization, Make- in-India, Save the girl child

1. INTRODUCTION

India is the largest producer of pulses, milk and the second largest producer of wheat, rice, sugarcane, groundnuts, fruits and vegetables (FA0, 2018). Agriculture in India contributes 18% of GDP and employed 50% of workforce in this sector (Financial Express, 2018). However, the average range of food loss/waste is approximate 4.6% -1 5.8% throughout the supply chain (India grows). This loss is mostly in the initial and mid stages of the food supply chain. Food losses (wastage) engender food insecurity, hunger, economic and environmental burden in the country (Global food loss). Moreover, it creates shortage of food to the children in India. Hence, India ranks 94 out of 107 countries and the Global Hunger Index (GHI) score of 27.2 in the year 2020 (Von Grebmer et al., 2020). Additionally, 40% of children are suffering from any type of malnourishment (approximate) and this will cause 11% GDP loss from malnutrition problems (Outlook Magazine, 2019).

To solve this alarming situation, Nutraceuticals "a food (or part of a food) that provides medical and/or health benefits, including the prevention and/or treatment of a disease" play a vital role in mitigating the risk of malnutrition in the world (Chaurasia *et al*., 2021). In addition, World Food Programme (WFP), UNICEF distributed a Nutraceutical Product named Plumpy Nut (RUTF, Ready-to-use Therapeutic food) in developing and under developing nations. Plumpy Nut is primarily patented and developed by French based Nutriset company. It is high energy, proteinaceous peanut based paste formulation for nutritional rehabilitation of children suffering from malnutrition from six months up to five years of age (Swaminathan, 2018; Guimón and Guimón, 2009). Their production facilities are spread over in Africa, France, USA, Latin America and India (Nutriset, 2019). However, 82% of all production facilities are in Africa countries as Ethiopia, Sudan, Burkina Faso, Niger, and Madagascar etc. (Parmigiani and RiveraSantos, 2015).

WFP procures only 75-80% of total demand for Plumpy Nut globally for malnourished children under Humanitarian supply chain. Therefore, 25-20% of the total demand is still unfulfilled. In addition, present industry is facing institutional voids such as regulatory ambiguities, labor, technical skill, contracting, product and capital markets due to Globalization (Khanna and Palepu, 1997; Khannaetal., 2005; Mairetal., 2012).

Parmigiani and Rivera-Santos (2015) emphasized the various risks associated with the global supply chain while providing Plumpy Nut under Public Private Partnership to severely malnourished children while coordinating between multiple partners. Hence, institutional voids create declining transparency and skepticism among customers, dietary guidelines issues, product recalls (Roth et al., 2008; Binns et al., 2018; Santini et al., 2018). Furthermore, melamine contamination in infant milk powder is a classic example for product recalls. It has reported 300,000 illnesses and six infants' death in the year 2008 (Roth et al., 2008; Marucheck et al. 2011). Hence, UNICEF implemented Sustainable Procurement in the year 2011 to fulfill total demand and reduce institutional voids for Plumpy Nut. This initiative has increased local markets from 0 to 48%, reduced cross continent trade from 90% to 60%, air freight from 10% to 0%, decrease in greenhouse gas emissions (GHG) of up to 89%, cost reduction up to 21% in the year 2017 (UNICEF Supply Chain Division, 2019). Thus, it has positive impact on economic, environment, social pillars of Sustainability.

In the same line of thought, India can also enter into the localization of Plumpy Nut. They can procure and processed local RUTF products for malnourished areas based on their malnutrition status (MOSPI, 2018). Therefore, the local Plumpy Nut has a cost of 2.60 €/kg whereas the global has an average cost of about 3 €/kg (Nutriset, 2010; UNICEF, 2009, food chemistry, 2013). Such approach has manifold impact especially in a price sensitive market like India.

Economically, cost is always a major criterion in today's society for decision and policymakers. Thus, affordability of the product will increase the number of malnourished children and improve the nutrition status of children from social perspective (Goal 2: Zero Hunger). This action can reduce the threat of environmental burden due to fewer transportations (Goal 12: Responsible Consumption and Production). The current study tries to demonstrate a holistic model of global and local supply chain while incorporating various stages of supply chain as procurement, production, processing, transportation and the final consumption. Furthermore, we investigate the expansion of local plant for solving the hunger and malnutrition challenges of India. In this context, we will try to answer through given research questions:

RQ1. *What are the various strategies to provide Plumpy Nut-Nutraceutical products for Indian malnourished children?*

RQ2. *What is the Optimal portfolio choice considering localization versus globalization of Plumpy Nut products?*

RQ3. *Is localization better than globalization of Plumpy Nut products with respect to the dimensions of Sustainability?*

RQ4. *How this research help policy makers, manufacturers and government in achieving UNSDG goals-2030 of SDG-2 (Zero Hunger) and SDG-12 (Responsible Consumption and Production)?*

Subsequent sections are arranged as follows: second section explains the nexus between Plumpy nut to solve malnutrition problems in developing/ under developing countries, operations research in solving nutritional diet problems, sustainable diet through literature review. Section three, we tried to describe the model over a given time period, section four discusses the analytical results. Furthermore, conclusion section describes the policymakers and stakeholders to take appropriate decisions for the eradication of malnutrition in their relevant countries.

2. LITERATURE REVIEW

2.1. Plumpy Nut for solving Malnutrition

Studies over the years highlighted the importance of good health and well-being corresponding to increase in economic output (Loeppke et al. 2007, World Bank 1993). Malnutrition aggravate the children's physical and mental growth, which further engender decrease in productivity and economic growth. Therefore, improving the child nutrition is vital to build the growth of the nation.

Several studies emphasized the impact of Plumpy Nut (RUTF) for curbing the situation of malnutrition in developing and under developing nations (Swaminathan, 2018; Scott-Smith, 2013; Ali et al., 2013; Nga et al., 2013). It has also been studied to provide nutrition to malnourished children in India (Swaminathan, 2018), Bangladesh (Ali et al., 2013), Vietnam (Phuong et al., 2014), Africa (Jayashankar and Swaminathan, 2014; Dery et al., 2018).

Guimón and Guimón (2012) studied diffusion of Plumpy Nut as an invention to fight hunger. They emphasized the expansion of local sources in affordable cost, less capital investment, a greater number of jobs for local people, organizational simplicity, overall acceptability based on their social environment. Another study by Santini et al. (2013) studied alternative Plumpy nut products with new product diffusion and formulations for malnourished children. Michaelsen et al. (2009) presented paper at WHO, WFP for the choice of food ingredients under five age for malnourished children. They described the characteristics of nutritious diets as high content of micronutrients, high energy density, adequate and high protein quality and high protein availability, adequate fat content, acceptable taste and texture, easy to prepare, affordable and accessible. In the similar line, Nutriset firm is also promoting for he localization of Plumpy Nut in their own countries and also invention of local available raw materials. Their teams are transforming peanut with other protein-based foods as Chickpeas, soy, corn, lentils to make product easily accepted by the targeted populations (Nutriset, 2019). Hence, it will reduce the seasonality of agriculture raw materials, contaminations, food quality and safety with reduced cost.

2.2. Operations Management in Nutrition/ diet problems

Nutrition/ diet problems are one of the oldest problems over the recent years. These studies indicated an emerging trend in solving diet problems with linear, goal, mixed integer linear, stochastic programming (Anderson et al., 1983; Lancaster (a), 1992; Lancaster (b), 1992; Dantzig, 1990; Briend et al., 2003; Mamat et al., 2012; Clerfeuille et al, 2013). However, there is still a rising concern among researchers, practitioners due to the growing population, current consumption patterns, perishability, food insecurity, climate change and environment concerns. Therefore, there is a growing concern of solving a holistic model in the context of sustainable supply chain design.

Sustainable Supply Chain design with respect to dietary considerations is of prime importance due to current production and consumption pattern, food security and health problems. Moreover, Sustainable diet is a vital part for everyday life. Hence, overconsumption of food, food waste should be reduced for our healthy diet and living status. Sustainable diet has three indicators such as i) Economic for affordable cost, ii) Environmental for climate change, arable land use, water use, fossil fuel depletion and iii) Social for improvement in dietary health (Rohmer et al., 2019). This article studied by using linear programming and Life Cycle

Assessment (LCA) methods to solve product mix problem in which small portion of beef has been replaced with spinach vegetables in the European country.

Another study by Peters et al (2016) focused of Nutrition Supply Chain with international, regional and local suppliers under World Food Programme (WFP). This paper studied optimization of food basket with local sourcing of food commodities and cash vouchers to solve long term humanitarian food aid problems with the help of Mixed Integer Linear Programming in natural disaster-prone countries (Iraq, Yemen) for helping three to six million beneficiaries. In this problem, rice is swapped with wheat, sorghum in the current food basket with reduction in cost and high in energy (Kilo Calories). According to WFP, ideally100% of recommended daily intake is necessary for the beneficiaries, but in this context 95% of required nutrient intake will reduce the cost of complete operation and increases 15-20% more beneficiaries in all over the world. Additionally, local procurement helps in reducing the cost and accessibility of nutritious products (Dibari et al., 2012; Komrska, 2012).

3. PROBLEM DESCRIPTION AND MATHEMATICAL MODEL FORMULATION

In this section, we tried to identified the given gaps in the literature review by designing the integrate model of Nutraceutical supply chain (NSC) for malnourished children. The detailed description of NSC is represented in Fig. 1. NFSC is a multi-echelon SC starting from the flow of raw materials (farmers), processing stages (nutraceutical industry), warehousing/ storing, followed by transportation through various modes, and finally the downstream distribution of finished products to the consumers based on their demand requirements. Costs related to each echelon of the NSC network design are emphasized in the model description.

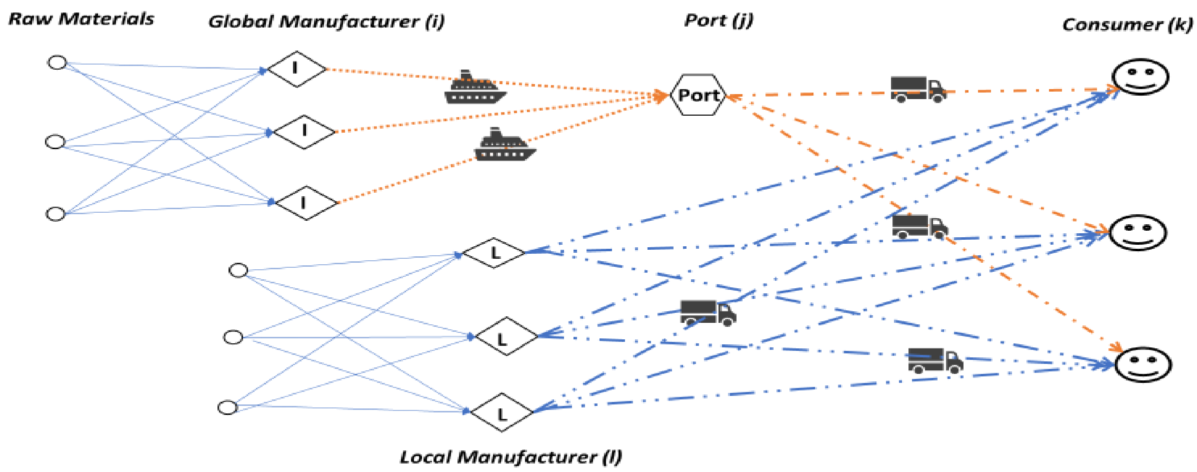


Figure 1**:** Network design for Globalization and Localization of Nutraceutical product Products

3.1. Mathematical Model formulation and assumptions

To solve the real complex problem of malnutrition in scarce resources situation, we have considered few assumptions to formulated the model:

- All parameters are deterministic in nature. Units of cost is considered as Rs. / ton in the model.

- Demand of each consumer locations are known and it's stable during a given period. We have assumed the number of malnourished children is stable means we are not considering child birth and death in the given period. Apart from this, we assumed only those malnourished children which are suffering from acute malnourishment stage. Children who are critically ill and towards the end phase of their life are not considered.
- Capacity of all manufacturing locations are already considered.
- Supply of raw material such as groundnuts/ peanuts are already known.
- Plumpy Nut can be distributed through both global and local manufacturers. Plumpy Nut has a shelf life of two years under normal atmospheric conditions. Hence, it doesn't require refrigeration type of facilities.
- Global manufacturers are sending processed food products through international airport and then to the consumer regions. Processed foods are warehoused/ stored in the port for 5-6 days for quality inspection and then it is dispatched. Therefore, we made assumptions that transportation mode outside India through ship (sea mode) whereas truck for road transportation inside India.
- Whereas, local manufacturers are sourcing raw materials (like peanuts, milk powder etc.) from the farmers. Peanuts are shelled and it have food loss which is represented as wastage in the model formulation. It is further processed with all raw materials in the local manufacturing plants and finally transported through trucks to the consumers.
- In the base scenario model, we have considered that there is no food loss and demand meets to feed all malnourished children in each region. Further, we portrayed to create shortage in the supply chain. Hence, shortage and wastage are added as binary variable in the supplement scenario model.
- Production and transportation lead time is not considered.
- Environment Cost of Groundnut/ Peanut is incorporated in the model as it is the major contributor of greenhouse gas emissions (GHG) as raw material in the model.

Based on the above assumptions, a Mixed integer linear programming (MILP) model is provided below:

Sets:

S Set of manufacturing locations ($S \subset I, R$)
I Set of global manufacturing locations ($i \subset I$)
R Set of local manufacturing locations ($r \subset R$)
L Set of farmer locations ($l \subset L$)
J Set of port locations ($j \subset J$)
K Set of consumers (malnourished children) locations ($k \subset K$)
M Set of transportation mode ($m \subset M$) (such as sea, truck)

Parameters:

C_i Cost of Plumpy Nut product (INR/ metric tonne) in global manufacturing location i
C_l Cost of Plumpy Nut product (INR/ metric tonne) in local manufacturing location r
T_{ijk} Cost of transportation (INR/km) from global manufacturing location i through port locations j to consumer locations k with transportation mode m
T_{lrk} Cost of transportation (INR/km) from farmers' locations l to the local manufacturing location r to consumer locations k with transportation mode m
d_{ijk} Distance (in km) from global manufacturing location i through port locations j to consumer locations k
d_{lrk} Distance (in km) travelled from farmers' locations l to the local manufacturing location r to consumer locations k
Cw_i Warehousing cost (INR/days) in global manufacturing location i
Cw_l Warehousing cost (INR/days) in local manufacturing location r

Cf_i Carbon footprint (tonnes CO_2 equivalent emissions/metric tonne of product) of Plumpy Nut product in global manufacturing location i

Cf_l Carbon footprint (tonnes CO_2 equivalent emissions/metric tonne of product) of Plumpy Nut product in local manufacturing location r

Cf_{ijk} Carbon footprint (tonnes CO_2 equivalent emissions/metric tonne/ km) due to transportation by ship from global manufacturing location i through port locations j to consumer locations k

Cf_{lrk} Carbon footprint (tonnes CO_2 equivalent emissions/metric tonne/ km) due to transportation by road from farmers' locations l to the local manufacturing location r to consumer locations k

Cfw_i Greenhouse gas emissions (tonnes CO_2 equivalent emissions/metric tonne of product) due to warehousing at port locations j from global manufacturing locations i

Cfw_l Greenhouse gas emissions (tonnes CO_2 equivalent emissions/metric tonne of product) from local manufacturing locations r

F_i Food Quality (dimensionless) of Plumpy Nut product in global manufacturing location i

F_l Food Quality of Plumpy Nut product in local manufacturing location r

F_{imc} Number of fed malnourished children due to global manufacturing location i at consumer locations k

F_{lmc} Number of fed malnourished children due to local manufacturing location r at consumer locations k

D Total demand (metric tonnes) of Plumpy Nut product at consumer locations k

PC_s Penalty Cost (INR/ metric tonne) due to shortage at consumer locations k

PC_w Penalty Cost (INR/ metric tonne) due to wastage at consumer locations k

D Total demand (metric tonnes) of Plumpy Nut product at consumer locations k

I Total plant capacity (metric tonnes) of global manufacturing location i

L Total plant capacity (metric tonnes) of local manufacturing location r

S_L Supply of raw material (metric tonnes) from farmer locations l

W Carrying capacity (metric tonnes) of port locations j

***Decision Variables*:**

Q_{ijk} Quantity (metric tonnes) of Plumpy Nut from global manufacturing location i through port locations j to consumer locations k

Q_{lrk} Quantity (metric tonnes) of Plumpy Nut produced at local manufacturing locations r to consumer locations k

α Acceptability score of local Plumpy Nut products (0-100%)

s Food shortage (metric tonnes) at consumer locations k (0-30%)

w Food wastage (metric tonnes) at consumer locations k (5-20%)

Objective Function

The objective function of the NSC model is to minimize the Total Cost for the network design. NSC network design is based on the three dimensions of Sustainability. Therefore, Total Cost comprises of Economic, Environment and Social Cost in the given model presented as Eq. (1)

$$minimize$$
$$\textbf{Total Cost } (\boldsymbol{TC}) = \text{Economic Cost}(EC) + \text{Environment Cost } (EnC) + \text{Social Cost } (SC) \quad (1)$$

<u>Economic side:</u>

The first dimension of Sustainability is Economic Cost, which is the combination of production/ processing, transportation and warehousing/ storages stages of NSC network design in Eq. (2). In addition, production and storage cost is the sum product of cost of finished product/ raw materials and the quantities supplied by global/ local manufacturers. Whereas, transportation cost is the sum product of the transportation cost per kilometer, distance covered between plants to the consumers and the quantities shown in Eq. (3)

Economic Cost(EC) = Production Cost(PC) +Transportation Cost (TC) +Storage Cost (SC) (2)

$$EC = \Bigg[\Bigg\{\sum_{i\in I}\Bigg(C_i * \sum_{j\in J}\sum_{k\in K} Q_{ijk}\Bigg) + \sum_{l\in L}\Bigg(C_l * \sum_{r\in R}\sum_{k\in K} Q_{lrk}\Bigg)\Bigg\} + \Bigg\{\Bigg(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} T_{ijk} * Q_{ijk} * d_{ijk}\Bigg) + \Bigg(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} T_{lrk} * Q_{lrk} * d_{lrk}\Bigg)\Bigg\} + \Bigg\{\sum_{i\in I}\Bigg(C_{wi} * \sum_{j\in J}\sum_{k\in K} Q_{ijk}\Bigg) + \sum_{l\in L}\Bigg(C_{wl} * \sum_{r\in R}\sum_{k\in K} Q_{lrk}\Bigg)\Bigg\}\Bigg] \quad (3)$$

Environmental side:

The specific and the second dimension is Environmental Cost is given below in Eq. (4). All costs related three stages of NSC network design is the sum product of carbon footprint due to processing/ transportation/ warehousing with the quantities provided by both manufacturers respectively.

Environmental Cost (EnC) = Environmental Cost of product p due to Processing (EI_P) + Environmental Cost of product due to Transportation (EI_T) + Environmental Cost of product due to Warehousing (EI_w) (4)

$$EnC = \Bigg[\Bigg\{\sum_{i\in I}\Bigg(Cf_i * \sum_{j\in J}\sum_{k\in K} Q_{ijk}\Bigg) + \sum_{l\in L}\Bigg(Cf_l * \sum_{r\in R}\sum_{k\in K} Q_{lrk}\Bigg)\Bigg\} + \Bigg\{\Bigg(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Cf_{ijk} * Q_{ijk}\Bigg) + \Bigg(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Cf_{lrk} * Q_{lrk}\Bigg)\Bigg\} + \Bigg\{\sum_{i\in I}\Bigg(Cfw_i * \sum_{j\in J}\sum_{k\in K} Q_{ijk}\Bigg) + \sum_{l\in L}\Bigg(Cfw_l * \sum_{r\in R}\sum_{k\in K} Q_{lrk}\Bigg)\Bigg\}\Bigg] \quad (5)$$

Social side:

The last and an important dimension of this network design is Social Cost which illustrated in Eq. (6). Further, it is addressing food quality (protein quality in this context), and number of children which is conspicuously feeding by the manufacturers in the Eq. (7-8). While all demand is met with the food supplied from manufacturers in the basic model. However, we have included food wastage and shortage as Eq. (9) and Eq. (10) as an important criterion when supply is not able to meet all requirements of demand side in the NSC model.

Social Cost (SC) = Food quality (FQ) + Number of Fed Malnourished Children (Fmc) + Food Shortage *(Fs)* + Food Wastage *(Fw)* (6)

$$FQ = \left[\left\{\sum_{i\in I}\left(F_i * \sum_{j\in J}\sum_{k\in K} Q_{ijk}\right) + \sum_{l\in L}\left(F_l * \sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\}\right] \tag{7}$$

Number of Fed Malnourished Children (Fmc)

$$Fmc = \left[\left\{\sum_{i\in I}\left(F_{imc} * \sum_{j\in J}\sum_{k\in K} Q_{ijk}\right) + \sum_{l\in L}\left(F_{lmc} * \sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\}\right] \tag{8}$$

Food shortage (Fs)

$$F_s = PC_s * s \tag{9}$$

Food wastage (Fw)

$$F_w = PC_w * w \tag{10}$$

Subject to the Constraints

Demand constraints

$$\left\{\left(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Q_{ijk}\right) + \left(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\} \geq D \tag{11}$$

$$\left\{\left(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Q_{ijk}\right)\right\} \geq (1-\alpha) * D \tag{12}$$

$$\left\{\left(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\} \geq \alpha * D \tag{13}$$

Plant Capacity constraints

$$\left\{\left(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Q_{ijk}\right)\right\} \leq I \tag{14}$$

$$\left\{\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Q_{lrk}\right\} \leq L \tag{15}$$

Supply constraints

$$\left\{\left(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\} \leq S_L \tag{16}$$

Port constraints

$$\left\{\left(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Q_{ijk}\right)\right\} \leq W \tag{17}$$

Shortage constraints

$$\left\{\left(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Q_{ijk}\right) + \left(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\} \geq (1-s) * D \tag{18}$$

Wastage constraints

$$\left\{\left(\sum_{i\in I}\sum_{j\in J}\sum_{k\in K} Q_{ijk}\right) + \left(\sum_{l\in L}\sum_{r\in R}\sum_{k\in K} Q_{lrk}\right)\right\} \leq (1-w) * D \tag{19}$$

Non-Negativity constraints

$$Q_{ijk} \geq 0 \ \forall\, i, j, k \tag{20}$$

$$Q_{lrk} \geq 0 \ \forall\, l, k \tag{21}$$

$$s + w = 1 \tag{22}$$

$$s \geq 0 \tag{23}$$

$$w \geq 0 \tag{24}$$

While optimizing objective function/s, we should consider constraints to solve the entire model. Therefore, Eq. (11-13) is subjected to demand constraints i.e. the processed foods supplied by global and local manufacturers should be greater than the total demand at each demand locations. In addition, α is the acceptability percentage for local processed foods, hence, (1- α) is illustrated for global processed foods in the Eq. (11), (12) respectively. Local and global processed foods should be less than the total demand from malnourished children in each node represented as Plant Capacity constraints under Eq. (14), Eq. (15). Both plant capacity constraints are found to be binding in this case study. Furthermore, raw materials are harvested and provided by farmers such as shelled peanuts. Hence, we have included shelled peanuts/ groundnuts in the Eq. (16). All finished products from global side is transported through sea ports. Hence, the quantities should be less than the given carrying capacity of all warehouses (Eq. 17) in this context.

From Eq. (11-16) are suggested under base scenarios of the model whereas Eq. (18) and Eq. (19) are studied under supplement scenarios as wastage and shortage constraints. Shortage (*s*) and wastage (*w)* portrayed as percentages which is subtracted from one and further multiplied it with the demand and it should be less than the processed foods from both manufacturers. Since shortage/wastage are binary variable, hence, both should be equal to one depicted in the Eq. (22). Eq. (20-24) are incorporated as non-negativity constraints of the MILP model.

Figure 2: Network design map view of the Global manufacturing plants

3.2. Case Study

To solve the real situation of fulfilling malnutrition to the malnourished children is highly complex and unrealistic in nature. Therefore, we considered a deterministic model of supplying nutrition with Plumpy Nut to the acute malnourished children suffering from malnutrition from six months up to five years of age in a period in the India. We also assumed the number of malnourished children is stable means we are not considering child birth and death in the given period. Moreover, Plumpy Nut can be distributed through global and local manufacturers. In addition, we have considered three supply regions from both manufacturers and three high malnourished states as demand centers in India. Global manufacturers (Africa, France, USA) are sending finished products through three international ports (Mumbai, Vizag, Chennai) to the consumer places (Jharkhand, Odisha, Karnataka) presented in the Fig. 2.

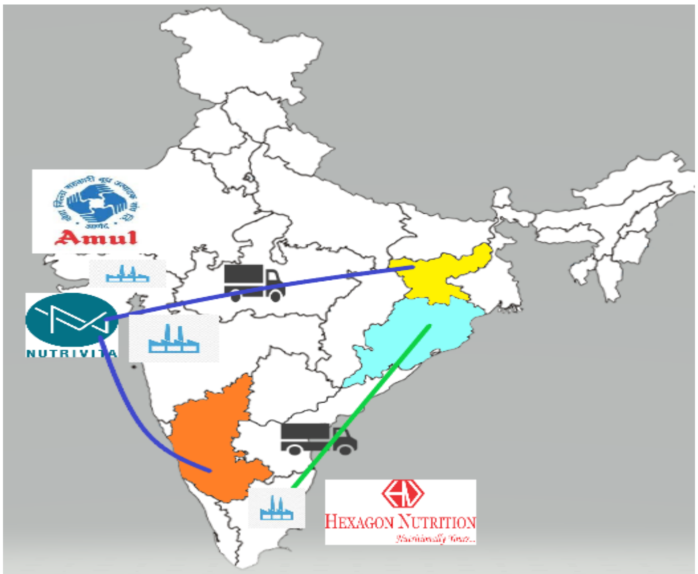


Figure 3: Network design map view of the Local manufacturing plants

Whereas, local manufacturers (Pune, Chennai, Gandhinagar) are procuring raw material from their own states (Maharashtra, Tamil Nadu and Gujarat) and further sending to their final destination consumer places illustrated in the Fig. 3. Since the current scenario is not coming under urgent situation. Therefore, we have assumed transportation mode outside India through ship (sea mode) whereas truck for road transportation inside India.

3.2.1. Data input

The supply data from global and local manufacturing plants are given in the appendix I under supply sources. Note all data and its sources are provided under Appendix I. Further, the number of malnourished children in India were retrieved from the Ministry of Statistics and Programme Implementation, Government of India (MOSPI, 2018). Though this report is based on National Family Health Survey NFHS-4 (2015-2016), therefore, we assumed the increase in the malnourished children data inputs for their respective regions based on the population growth of India, that is further added to the database since year 2016.

According to UNICEF Supply chain division, one carton of Plumpy Nut contains 150 sachets of 92 gm and it is sufficient for the treatment of one child with 10-15 kg over a period of 6-8 weeks. Therefore, we have considered that 10 kg of Plumpy Nut is essential to treat one child within two months of period (UNICEF, 2021). Hence, total number of malnourished children multiplied by 10 factor provides us the total demand from the respective regions of India.

Previous section 3.1 explained the components for Economic cost as production cost, transportation cost and warehousing cost. Thus, we have extracted production cost for finished products from the UNICEF prices for RUTF and case study published by Jarrod & Swaminathan (2015). However, raw material costs are obtained from online platform as Alibaba.com. Two different modes of transportation as ship for seaways and truck for roadways are considered in the study. Moreover, the distance between various countries are location wise and transportation mode dependent. We assumed that agricultural lands and the production plants are in close proximity to avoid post-harvest losses, lead time and improved customer service level. In addition to this, sea transportation costs have taken from DHL sources and truck transportation costs are selected from Indian database (ICCT, 2017). Note we have considered general cargo ship and

N3 rigid vehicle truck of 4x2 axle having speed 40 km/hr. for roadways transportation. Warehousing cost and carrying capacity of the India ports are extracted from the Ministry of Ports.

While calculating the environment cost due to production, transportation and warehousing cost, we considered 50$ per ton of GHG emissions. GHG emissions of peanut is 0.60, 0.90, 1.30, and 1.60 kg CO2 equivalent/ kg groundnut of USA, France, Africa and India respectively (McCarty *et al.*, 2016; Volpe *et al.*, 2015; Tongwane *et al.*, 2016; Jain *et al.*, 2016). Moreover, GHG emissions due to transportation is calculated through NTM Calc Basic 4.0 tool. This tool is easy to calculate the distance between two regions from various transportation modes with COx, NOx available database all over the world. Carbon emissions due to warehousing is almost negligible as Plumpy Nut doesn't require refrigeration methods. It has a shelf life of two years under ambient conditions.

Finally, we included social cost as an important component of total cost. In this, food quality is considered as protein percentage in the peanuts. Surprisingly, we found to notice that true social cost of malnutrition is 30$. This means that every 1 $ spent on child's nutrition will revert back as 30$ in the future (Hoddinott, 2013). Hence, the shortage cost associated with the Plumpy Nut to the malnourished children, will reduce 30 times in the near future. Note, for every calculation we have considered 74.5 Rupees for the Dollar conversion in the present study. Apart from these, Pluses such as Peanuts/ Groundnuts contribute 6.36 to 8.41% of food waste annually and is retrieved from the Indian data. The penalty cost due to food wastage information is extracted from the FAO (2013) database.

Though, NSC is highly complex as it includes a large number of objectives, variables and constraints. Hence, we tried to make our model holistic in nature because it includes producers, manufacturers, distributors, and consumers in mode. Further, MILP analysis is carried out in GAMS (Generalized Algebraic Modelling System) interface system (using CPLEX optimization tool)., which is a high-level modelling system. It helps in solving mathematical optimization as linear, nonlinear, and stochastic dynamic programming optimization problems. It can create an integrated development environment (IDE) and is connected to a group of third-party optimization solvers like CPLEX, PYTHON, DEA, XPRESS, COIN-OR solvers. Moreover, it has GAMS World Forum Digest that helps in discussing and solving modelling errors. Hence, it is fast, convenient, advanced, and user-friendly to integrate various tools under one platform.

4. COMPUTATIONAL RESULTS AND DISCUSSSIONS

In this following section, we validated the significance of present research illustrated in previous sections. In addition, the model is tested for different scenarios based on the mathematical formulation and the data inputs described/used in the section 3.1 and 3.2 respectively. As a result, the model is analyzed for various scenarios based on the aforementioned economic, environment, and social objectives for minimizing the total cost as a single objective. Therefore, different optimal solutions for global and local Plumy Nut are interrelated to each other. Note, that all costs and their associated values are considered as Rs/ton in the vertical axis of the graphs. While doing analysis, we tried to explain the findings of total cost estimation for global and local scenarios are depicted in the Fig. 4.

As discussed above, the total cost function is the summation of economic, environment and social cost. In Fig. 4, the analysis has done on the localization of Plumpy Nut from 0 to 100% and further expansion of local pant capacity again from 0 to 100% (100-200% in the graph). Note, all values are given in percentages of the total demand demonstrated in the horizontal axis of the graphs further used.

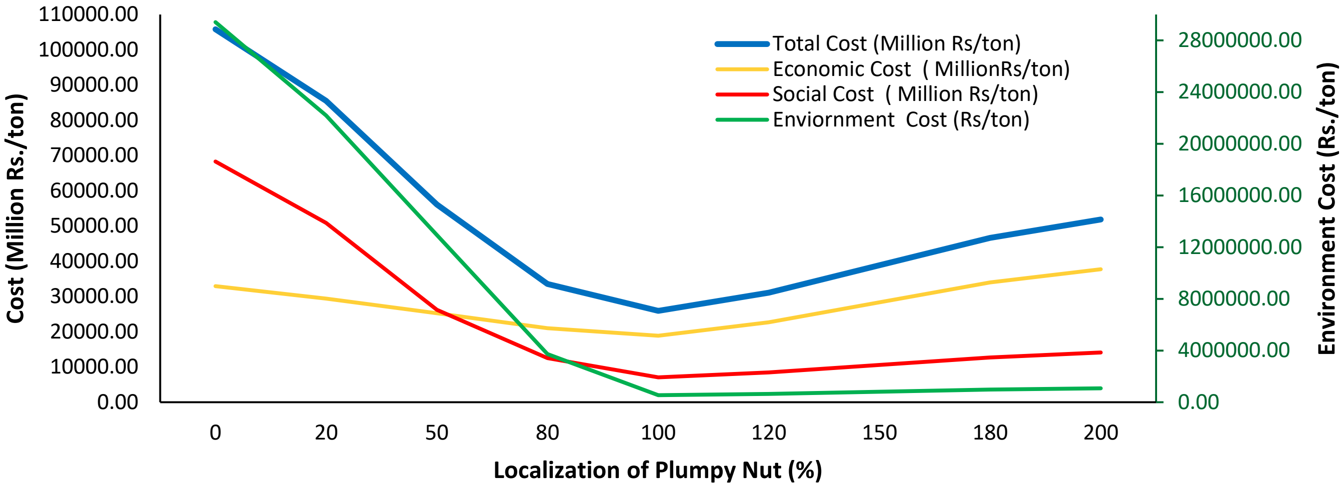


Figure 4: Total Cost esimation for Global and Local Plumpy Nut

.

In this context, Fig. 4 presents the total cost estimation for global and local Plumpy Nut. It is a slight 'U' curve function and with a steep decline while going from global to local, and a further slight rise when going for expansion of upto 200% for local products. A close look at the graph explained that social cost has a major contribution to the total cost, followed by economic cost. Table 1 explains a reduction of 76% (approximately) for total cost, when there is an increae in the total demand of Plumpy Nut from local sources instead of global ones for 7,446,362 malnourished children. Furthermore, if there is an increase in the the demand of up to 200%, it reduced the cost of Plumpy Nut up to 51% for 15,970,213 children. Hence, localization is a better option for providing nutrition through Plumpy Nut to the malnourished children instead of globalization.

Table 1: Costs used in the Plumpy Nut network design w.r.t. to the localization

Localization	Total Cost (%)	Economic Cost (%)	Environment Cost (%)	Social Cost (%)	Fed Malnourished Children
0-100	75.51	42.77	98.15	89.67	7446362
0-200	51.02	14.47	96.31	79.34	15970213

4.1. Economic side:

In connection with the total cost, economic cost is an important indicator of Sustainability as the best optimal solution depends on the decision makers perspective for global or local products under various scenarios. Since, allocation of Plumpy Nut to the given demand regions is highly complex as it depends upon the donor funding. Therefore, economic cost is an important consideration especially from the manufacturers and the funding agencies side. It is a critical parameter as the best optimal solution depends on the decision makers perspective for global or local products under various scenarios. Fig 5 is a 'U' shape graphical presentation for the economic cost. While going from global to local Plumpy Nut, 43% reduction in the economic cost illustrated in the Table 2. The present graph (Fig. 5) shows a sharp decline and increase at more extreme points for global and local sources, while the kink at the middle of the graph (*i.e.* 100% localization) illustrates a better optimal solution.

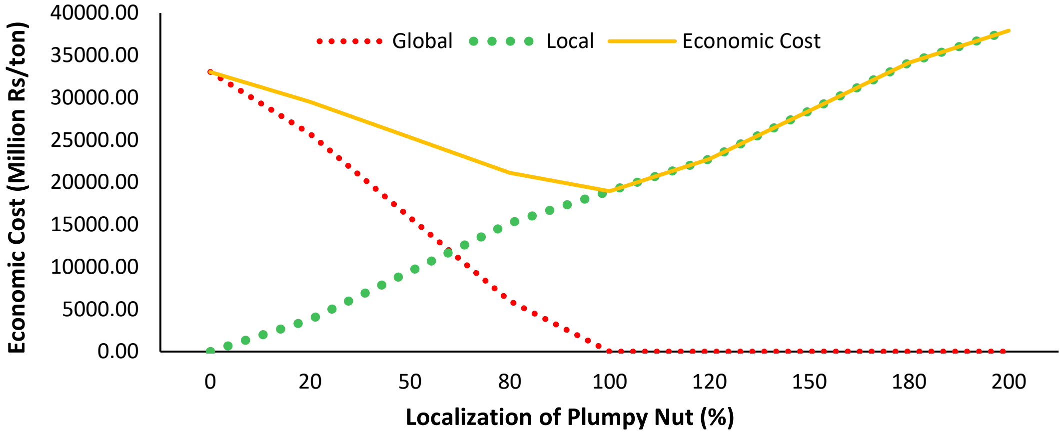


Figure 5: Economic Cost for Global and Local Plumpy Nut

.
Furthermore, the red dotted line shows a steep decrease whereas green dotted line shows a steep increase in the linear trend. Both red and green dotted line intersects each other between the range of 50-80% of localization.

In the previous section, we have considered 7446362 as the total number of malnourished children from the demand regions. Thus, 0% localization (100% globalization) costs 4426 rupees to treat one malnourished child. In addition, 20, 50, 80, 100 to 200 percentage of localization costs around 3952, 3388, 2825, 2533 rupees to treat one malnourished child. Thus, the analysis investigate that decision makers should adopt localization from 0 to 100 percentage, however, beyond that point, it will further increase the cost of the treatment due to plant capacity expansion and other variable costs. From economic point of view, if supply from local plants met the demand requirements, then decisionmakers should opt for to 100% local products otherwise further expansion would be helpful in fulfilling the total demand for malnourished children in this context.

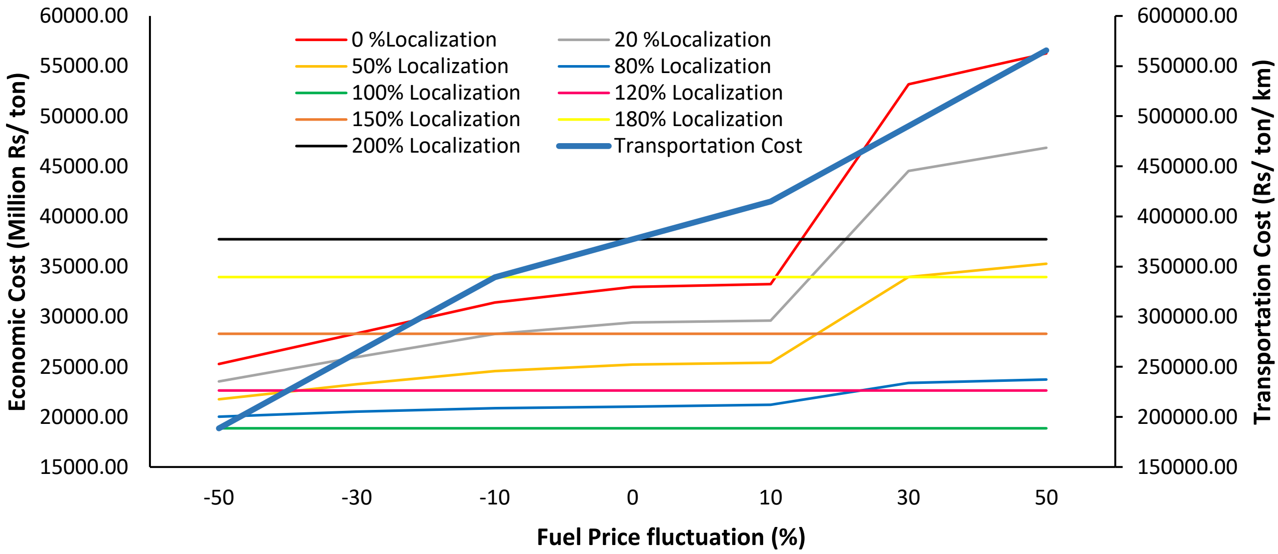


Figure 6: Economic Cost due to fluctutation in Transportation Cost

Though, economic cost consists of raw material, transportation and warehousing cost as explained in the Eq. 3, 4 under section 3.1, transportation cost plays a prominent role in deciding the economic cost. It fluctuates easily due to the uncertain fuel price dynamics in the market, therefore, transportation impacts the complete supply chain. We have done sensitivity analysis to see the impact of fuel changes in the transportation and economic cost. Fig. 6 shows a linear line of transportation cost, when we fluctuate the fuel price upto 50%. Although, 0, 20 and 50, 80% localization portrays almost stable line while fluctuating the fuel price from -10% to 10% values. However, 10% increase in the fuel price creates a sudden rise in the economic cost for the global foods. Therefore, the logistics providers should not opt for global ones when the prices hike more than 10%. Apart from these insights, 100% to 200% of localization represents a flat line while fluctutating the fuel price from -50 to 50% values. Hence, it explains that transportation cost has negilgible impact in he econmomic cost for local ones. Finally, we could expalin that economic cost due to global depends upon raw material, transportation cost whereas local depends upon transportation cost only in this scenario. In addition, warehousing cost in insignificant in the global whereas found lacking in the local ones.

4.2. Environmental side:

Interestingly, global warming, increase GHG emissions fascinated the attention among environmentalists. Therefore, it is an imperative indicator among policymakers. This side of sustainability also includes environment cost due to the raw material/processing, transportation and warehousing given in the Eq.5, 6 under section 3.1. Raw material has a minor role in calculating the environment cost. However, Indian peanuts has the highest GHG emissions which is further followed by Ethiopian, French, and USA peanuts. While considering environment side, transportation is the most noticeable feature of the supply chain. It contributes 14% of total GHG emissions at the global level (EEA, 2011). Fig. 4 illustrates a curve in this objective function, starts with a steep decline which is followed into a flat continuation. In the steep decline part of the curve, upgraded environment cost can be obtained by relatively a small change in the localization of Plumpy Nut. Moreover, we have noticed 98% decrease in GHG emissions from global to local (first part of the curve), but 2% rise in GHG emissions for local expansion of up to 100% (flat part of the curve) given in the Table 2.

4.3. Social side:

Providing nutrition or improvement in dietary health is a case of Social dimension. Hence, Social cost includes a major input to the total cost as illustrated in the Fig. 4. Moreover, Eq. 7-9 explains the indicators of Social cost as food quality, number of fed malnourished children. In this case study, protein is considered as food quality dimension. However, protein quality is better in global sources as compared to local sources. This means peanuts from USA, France and Ethiopia has high percentage of protein as compared to the Indian peanuts. Therefore, Indian peanuts should have high percentage in the Plumpy Nut in order to fulfill the basic protein diet requirements for the malnourished children.

Furthermore, the number of fed malnourished children with respect to the localization of Plumpy Nut is presented in the Fig. 7. Note we have assumed 66,220,000 (GHI = 27.2) Indian malnourished children in the present year as data input under the section 3.2. For this, we have considered 7,446,362 children from all demand regions in the initial stage of the calculation. Furthermore, the graph explains that increase in the localization of Plumpy Nut decreases the social cost. Therefore, reduction in the social cost helps in increasing the number of fed malnourished children depicted as green line in the provided figure.

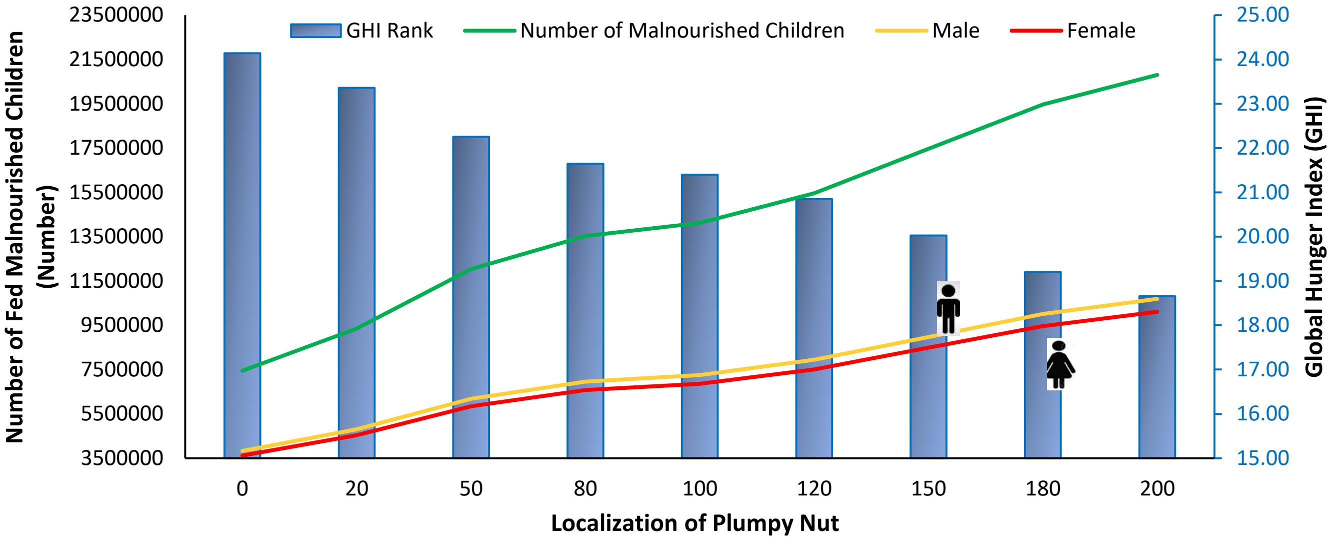


Figure 7: GHI and Fed Malnourished Children for Global and Local Plumpy Nut

Such as 20% localization of Plumpy Nut may reduce the social cost of up to 10.7%. Thus, if we add 10.7% to the current malnourished children which further increases the number of malnourished children in this step as 8,243,123. Moreover, if we subtract the fed children (8,243,123) from the total assumed unfed children (66,220,000), it gives us the final unfed number of malnourished children (57,976,877). Hence, the final unfed number of malnourished children maintains the GHI score of 23.81 and it is portrayed as bar diagram in the Fig. 7.

Female child is biologically stronger than male child as nature provides strong immunity in terms of 'XX' chromosomes in female as compared to the male ones who is the recipient of single 'X' chromosomes. In addition, all over the world male mortality rate due to various diseases/ malnutrition is high with respect to the female ones whereas in India, it is vice versa. However, we discovered that male children get better nutrition than female children in India (presented as Yellow and Red line in the Fig.7). This might be that parents usually take care of their male child in terms of better nutrition, breast feeding, vaccination and oral supplementation due to cultural/ social issues (UNICEF, 2020). Even in the prosperous agricultural lands of Haryana and Punjab,

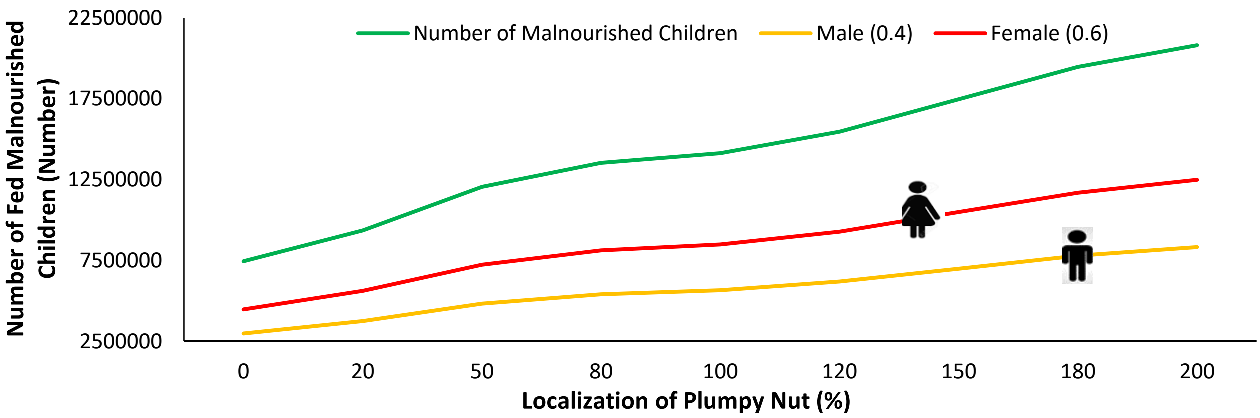


Figure 8: Male and Female Fed Malnourished Children for Global and Local Plumpy Nut

the situation is severe owing to the neglected behavior towards female infants (Khosla, 1980). Hence, Indian government implemented several meal schemes under *Integrated Child Development Services (ICDS)* and *Anganwadi* centers in the districts. These schemes help in promoting the slogan as *"Save the Girl Child" (Beti bachao, beti padao).* In addition, if we provide 60% of Plumpy Nut as compared to the male children as explained in the Fig. 8. It could help in reducing the root level problem of preferring/providing nutrition to the male children as compared to the female children.

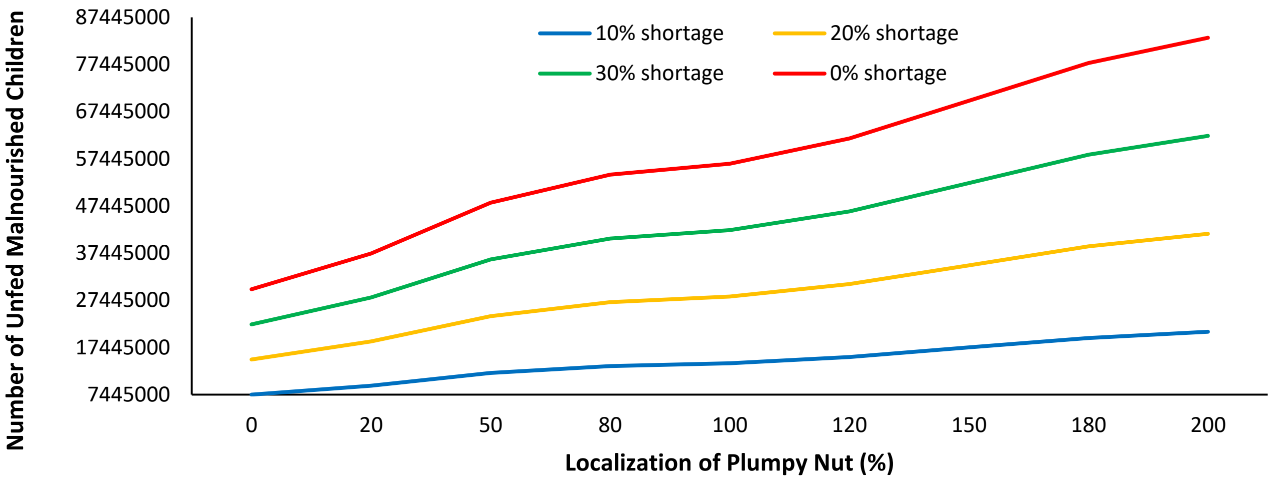


Figure 9: Unfed Malnourished Children due to Shortage for Global and Local Plumpy Nut

Needless to say, the supply of Plumpy Nut meets the total demand from all malnourished regions of India in the above scenarios. Therefore, there is no shortage and wastage in the NSC network design. However, we are trying to solve the uncertain demand of Plumpy Nut due to various environments. Therefore, we have included shortage and wastage cost under social dimension in the Eq. 10-11. When the supply of Plumpy Nut does not meet the total demand, it creates a shortage in the supply chain as presented in the Fig. 9.

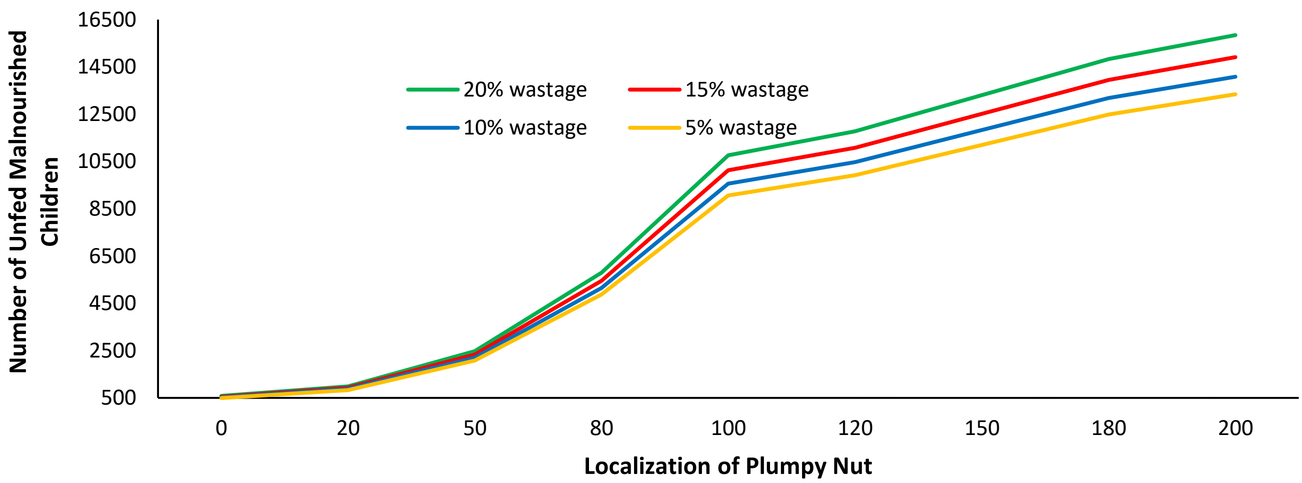


Figure 10: Unfed Malnourished Children due to Wastage for Global and Local Plumpy Nut

We have considered 10-30% of shortage in the NSC network design. As the shortage of Plumpy Nut increases in the demand regions, further it increases the number of unfed malnourished children in their regions also. In addition, Jharkhand has the highest number of malnourished regions, should focus first to the highest demand regions as compared to other regions of India. Another scenario is related to the wastage in the NSC network design illustrated in the Fig.10. Since, we are procuring shelled peanuts as raw material from the farmers. Therefore, it contributes 5% to 20% wastage during transportation and processing stages. As the wastage of Plumpy Nut increases in the demand regions, further it increases the number of unfed malnourished children in their regions also. Therefore, decisionmakers should focus first to the highest demand regions as compared to other regions of India.

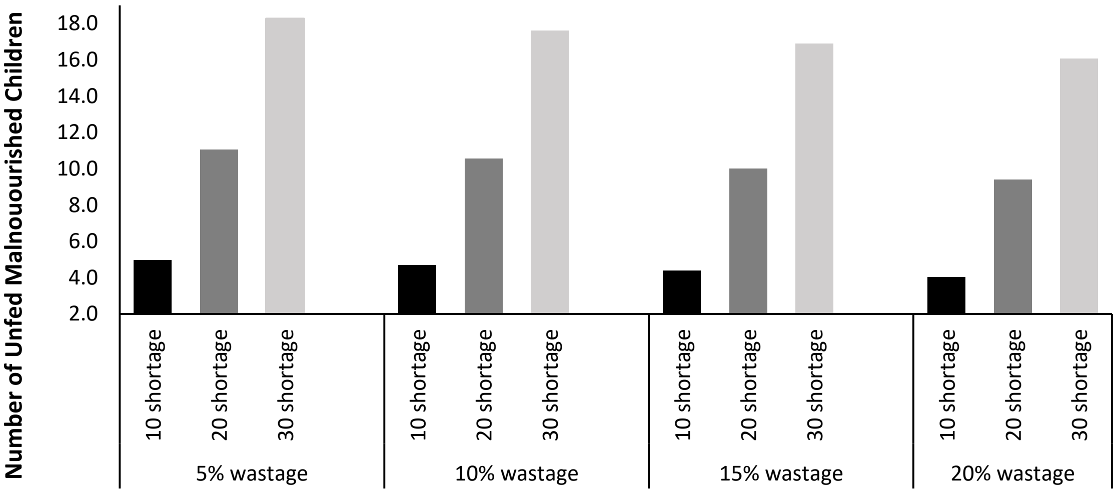


Figure 11: Effect of Wastage for Shortage of Plumpy Nut for Malnourished Children

Furthermore, if we reduce 5% wastage in the supply chain given in the Fig.11, provides 5% nutrition to the 10% shortage of the unfed malnourished children in NSC. Same as 5% wastage provides nutrition to the 11%, 18.3% to the 20 and 30% shortage of unfed children respectively. In addition, 20% wastage provides 4%, 9.4% and 16.1% of nutrition to the 10, 20 and 30% shortage of malnourished children respectively. Therefore, we suggest that less wastage helps in reducing the shortage of Plumpy Nut in the entire supply chain, which further helps in achieving the targets of SDG-12: *Responsible Consumption and Production goals*. Hence, this action will make decision makers to set coercive pressures on manufactures and logistics providers to reduce the waste in the NSC network design.

5. CONCLUSION

India is a land of agricultural resources as well as it is the land of malnourished children also. Additionally, it has the highest mortality rate of female children under the age if five in the world. This paper proposed a MILP model for NSC network design. It helps decision makers to adopt local Plumpy Nut food products for malnourished children. We included managerial insights and policymakers' implications based on our analytic findings in Table 2. Further, it elaborated the expansion of local capacity of the plants of up to 100%, which will increase the number of fed malnourished children in the future. This initiative will reduce the affordable cost (40%) and

Research Questions	Research Findings	Managerial and Policymakers implications
RQ1 & RQ2	(i)Local and global manufacturers may help in providing nutrition through Plumpy Nut to the Indian malnourished children. (ii) Analysis explains that Local manufacturer is better than global manufacturer. (iii) It produces affordable, easy to access, less GHG emissions, less lead time and locally accepted processed food products. (iv) Another interesting fact is that India has enough resources available but due to insufficient warehouses it leads to food waste and losses. (v) Reduction in RUTF cost and 95% nutrition available according to Pareto Optimal path reduces the number of malnourished children.	•This network design helps policymakers to take sustainable decisions when we go for local or global manufacturers depending on the natural resources available with desirable quality, cost, safe and nutritious food. • Localization will help ‘Make in India’ progress in his context. • This will further go hand in hand with Sustainability and Food Security perspective.
RQ3	(i) Economic indicates affordable cost, income growth for local workers, less food waste, less transportation and increased accessibility of foods; (ii) Environment involves less carbon footprint, fertilizers, pesticides use, water, biodiversity with availability, increased food production trade; (iii) Social includes proportion of meals consumed, acceptance of traditional products, high nutrition quality score, global nutrition index, security of vulnerable groups, women’s literacy to improve Utilization and stability in terms of variability of food prices, due to climate change, stable prices presented in the Fig. 4 (refer Appendix 2).	• Localization is better than globalization such as 42% w.r.t Economic cost, 98% w.r.t. Environment cost, 89% increase in the Social Cost dimension. • It will help in reducing the GHI of India. •Warehouses in India should be increased and properly managed with assured standards.
RQ4	(i)Female infants have a high under-five mortality rate exceeded than male infants under Social dimensions in Fig, 5, 6 (refer Appendix 2). (ii) Food wastage will help in reducing the shortage of foods in Fig.7. (ii) Peanut in Plumy Nut can be replaced with chickpeas, mung bean, pigeon pea, soy and lentils. They have less GHG emissions than Peanuts. (iv) Sustainable Innovation (diet mix) helps in minimizing the cost, less GHG emissions and increased availability of local raw materials.	• Government should focus on nutrition policies especially for female children under ‘Save the Girl Child’ slogan. • This will help them in Policymaking to achieve ***Zero Hunger*** and ***Responsible Consumption and Production*** under Public-Private Partnerships.

Table 2: Computational Analysis and Discussions

increase in the cost by approximately 15% in order to provide nutrition to the twice times number of fed malnourished children. Though, localization of Plumpy Nut reduced the GHG emissions due to few transportations of approximate 95%, but further expansion will contribute minor significant changes in the GHG emissions. Furthermore, if government will seek special attention to the female children. It will improve the GHI of India, which will further help in improving the overall nutrition to the child from his/her mother in this context. Apart from these implications if we reduced the food wastage, it helped in reducing the shortage of food for malnourished children in India. Hence, the present paper will help decision/ policymakers to implement and also achieve SDG 2: *Zero Hunger* and SDG 12: *Responsible Consumption and Production* in their developing/ under developing country.

REFERENCES:

Anderson, A.M. & Earle, M.D. (1983). Diet planning in the third world by linear and goal programming. *Journal of the Operational Research Society*, *34*(1), 9-16.

Ali, E., Zachariah, R., Dahmane, A., Van den Boogaard, W., Shams, Z., Akter, T., Alders, P., Manzi, M., Allaouna, M., Draguez, B. & Delchevalerie, P. (2013). Peanut-based ready-to-use therapeutic food: acceptability among malnourished children and community workers in Bangladesh. *Public health action*, *3*(2), 128-135.

Asibuo, J. Y., Akromah, R., Safo-Kantanka, O., Adu-Dapaah, H. K., Ohemeng-Dapaah, S., & Agyeman, A. (2008). Chemical composition of groundnut, Arachis hypogaea (L) landraces. *African Journal of Biotechnology*, *7*(13).

Bagchi, D. ed. (2019). *Nutraceutical and functional food regulations in the United States and around the world*. Academic press.

Bazzano, A.N., Potts, K.S., Bazzano, L.A. & Mason, J.B. (2017). The life course implications of ready to use therapeutic food for children in low-income countries. *International journal of environmental research and public health*, *14*(4), 403.

Berry, E.M., Dernini, S., Burlingame, B., Meybeck, A. & Conforti, P. (2015). Food security and sustainability: can one exist without the other? *Public health nutrition*, *18*(13), 2293-2302.

Briend, A., Darmon, N., Ferguson, E. & Erhardt, J.G. (2003). Linear programming: a mathematical tool for analyzing and optimizing children's diets during the complementary feeding period. *Journal of pediatric gastroenterology and nutrition*, *36*(1), 12-22.

Business Today (2021). https://www.businesstoday.in/current/economy-politics/india-ranks-94th-in-global-hunger-index-falls-into-serious-category-under-5-mortality-rate/story/4 (accessed on 4, June, 2021)

Carrillo-Hermosilla, J., Del Río, P. & Könnölä, T. (2010). Diversity of eco-innovations: Reflections from selected case studies. *Journal of cleaner production*, *18*(10-11), 1073-1083.

Chaurasia, S., Pati, R. K., Padhi, S. S., Jensen, J. M., & Gavirneni, N. (2021). Achieving the United Nations Sustainable Development Goals-2030 through the nutraceutical industry: A review of managerial research and the role of operations management. *Decision Sciences*.

Clapp (2015). FAO, 2015. http://www.fao.org/3/a-i5160e.pdf. (accessed on 30, Oct, 2019).

Clark, T. & Charter, M. (2007). Sustainable innovation: Key conclusions from sustainable innovation conferences 2003–2006 organized by the center for sustainable design.

Clerfeuille, E., Vieux, F., Lluch, A., Darmon, N. & Rolf-Pedersen, N. (2013). Assessing the construct validity of five nutrient profiling systems using diet modeling with linear programming. *European journal of clinical nutrition*, *67*(9), 1003.

Dantzig, G.B. (1990). The diet problem. *Interfaces*, *20*(4), 43-47.

De Pee, S. & Bloem, M.W. (2009). Current and potential role of specially formulated foods and food supplements for preventing malnutrition among 6-to 23-month-old children and for treating moderate malnutrition among 6-to 59-month-old children. *Food and nutrition bulletin*, *30*(3_suppl3), S434-S463.

Dekker, R., Bloemhof, J., & Mallidis, I. (2012). Operations Research for green logistics–An overview of aspects, issues, contributions and challenges. *European journal of operational research*, *219*(3), 671-679.

Dibari, F., Diop, E.H.I., Collins, S. & Seal, A. (2012). Low-cost, ready-to-use therapeutic foods can be designed using locally available commodities with the aid of linear programming. *The Journal of nutrition*, *142*(5), 955-961.

Duong, M. H. (2009). What is the Price of Carbon? Five definitions. *SAPI EN. S. Surveys and Perspectives Integrating Environment and Society*, (2.1).

Ehrgott, M. (2006). A discussion of scalarization techniques for multiple objective integer programming. *Annals of Operations Research*, *147*(1), 343-360.

FAO, IFAD, UNICEF, WFP and WHO (2018). The State of Food Security and Nutrition in the World 2018.

FAO (2013). Footprint, F.F.W., Impacts on Natural Resources; FAO: Rome, Italy.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO), 2014, Building a common vision for sustainable food and agriculture: Principles and Approaches, E-ISBN 978-92-5-108472-4, http://www.fao.org/3/i3940e/i3940e.pdf (assessed 28 September, 2021).

Financial Express (2018). https://www.financialexpress.com/budget/india-economic-survey-2018-for-farmers-agriculture-gdp-msp/1034266/ (accessed on 20, Oct, 2019).

Global Nutrition report (2020), https://globalnutritionreport.org/resources/nutrition-profiles/asia/southern-asia/india/ (accessed on 20, Nov, 2020).

Guimón, J. & Guimón, P. (2012). How ready-to-use therapeutic food shapes a new technological regime to treat child malnutrition. *Technological Forecasting and Social Change*, *79*(7), 1319-1327.

Haimes, Y. (1971). On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE transactions on systems, man, and cybernetics*, *1*(3), 296-297.

Hoddinott, J. (2013). The economic cost of malnutrition. *The Road to Good Nutrition* (64-73). Karger Publishers.

Hwang, C. L., & Masud, A. S. M. (2012). *Multiple objective decision making—methods and applications: a state-of-the-art survey* (Vol. 164). Springer Science & Business Media.

India Today (2021a). https://www.indiatoday.in/india/story/global-hunger-index-2020-india-ranks-94-1732452-2020-10-17 (accessed on 14 December, 2020).

India Today (2021b). https://www.indiatoday.in/india/story/india-grows-more-food-wastes-more-while-more-go-hungry-1752107-2020-12-22 (accessed on 10 Feb, 2021).

International Food Policy Research Institute. (2015). Global Nutrition Report 2015: Actions and accountability to advance nutrition and sustainable development. Washington, DC.

Jain, N., Arora, P., Tomer, R., Mishra, S.V., Bhatia, A., Pathak, H., Chakraborty, D., Kumar, V., Dubey, D.S., Harit, R.C. & Singh, J.P. (2016). Greenhouse gases emission from soils under major crops in Northwest India. *Science of the Total Environment*, *542*, 551-561.

Jarrod, G. & Swaminathan, J.M. (2015). UNICEF RUTF Supply Chain (B): Developing a Global Supply Network, University of North Carolina's Kenan---Flagler Business School.

Khosla, T. (1980). The plight of female infants in India. *Journal of Epidemiology & Community Health*, *34*(2), 143-146.

King, J. C., Blumberg, J., Ingwersen, L., Jenab, M., & Tucker, K. L. (2008). Tree nuts and peanuts as components of a healthy diet. *The Journal of nutrition*, *138*(9), 1736S-1740S.

Kovács, G., & Spens, K. M. (2011). Trends and developments in humanitarian logistics–a gap analysis. *International journal of physical distribution & logistics management*.

Kretschmer, A., Spinler, S., & Van Wassenhove, L. N. (2014). A school feeding supply chain framework: Critical factors for sustainable program design. *Production and Operations Management*, *23*(6), 990-1001.

Lancaster, L.M. (1992a). The evolution of the diet model in managing food systems. *Interfaces*, *22*(5), 59-68.

Lancaster, L.M. (1992b). The history of the application of mathematical programming to menu planning. *European Journal of Operational Research*, *57*(3), 339-347.

Layton, J. (1998). How Calories Work. *Retrieved July*, *15*, 2009.

Loeppke, R., Taitel, M., Richling, D., Parry, T., Kessler, R. C., Hymel, P. & Konicki, D. (2007). Health and productivity as a business strategy. *Journal of Occupational and Environmental Medicine, 49(7),* 712-721.

Longvah, T., Anantan, I., Bhaskarachary, K., Venkaiah, K., & Longvah, T. (2017). *Indian food composition tables* (pp. 2-58). Hyderabad: National Institute of Nutrition, Indian Council of Medical Research.

Mamat, M., Zulkifli, N.F., Deraman, S.K. & Noor, N.M.M. (2012). Fuzzy linear programming approach in balance diet planning for eating disorder and disease-related lifestyle. *Applied Mathematical Sciences*, *6*(103), 5109-5118.

Marucheck, A., Greis, N., Mena, C. & Cai, L. (2011). Product safety and security in the global supply chain: Issues, challenges and research opportunities. *Journal of Operations Management*, *29*(7-8), 707-720.

Mavrotas, G. (2009). Effective implementation of the ε-constraint method in multi-objective mathematical programming problems. *Applied mathematics and computation*, *213*(2), 455-465.

McCarty, J.A., Ramsey, S. & Sandefur, H.N. (2016). A historical analysis of the environmental footprint of peanut production in the United States from 1980 to 2014. *Peanut Science*, *43*(2), 157.

Michaelsen, K.F., Hoppe, C., Roos, N., Kaestel, P., Stougaard, M., Lauritzen, L., Mølgaard, C., Girma, T. & Friis, H. (2009). Choice of foods and ingredients for moderately malnourished children 6 months to 5 years of age. *Food and nutrition bulletin*, *30*(3_suppl3), S343-S404.

MOSPI(2018).http://www.mospi.gov.in/sites/default/files/publication_reports/Children%20in%20India%202018%20%E2%80%93%20A%20Statistical%20Appraisal_26oct18.pdf (accessed on June, 2019).

Natarajan, K.V. & Swaminathan, J.M. (2014). Inventory management in humanitarian operations: Impact of amount, schedule, and uncertainty in funding. *Manufacturing & Service Operations Management*, *16*(4), 595-603.

Natarajan, K.V. & Swaminathan, J.M. (2017). Multi-treatment inventory allocation in humanitarian health settings under funding constraints. *Production and Operations Management*, *26*(6), 1015-1034.

Nutriset (2019). https://www.nutriset.fr/en/plumpyfield (accessed on 25, Oct, 2019).

Outlook (2019). https://www.outlookindia.com/website/story/india-news-indias-hunger-index-what-can-be-done/326676 (accessed on 20, Oct, 2019).

Parmigiani, A. & Rivera-Santos, M. (2015). Sourcing for the base of the pyramid: Constructing supply chains to address voids in subsistence markets. *Journal of Operations Management*, *33*, 60-70.

Peters, K., Fleuren, H., Den Hertog, D., Kavelj, M., Silva, S., Goncalves, R., Ergun, O. & Soldner, M., (2016). The nutritious supply chain: Optimizing humanitarian food aid.

Phuong, H., Nga, T.T., Mathisen, R., Nguyen, M., Hop, L.T., Hoa, D.T.B., Minh, H.N., Tuyen, L.D., Berger, J. & Wieringa, F.T. (2014). Development and implementation of a locally produced ready-to-use therapeutic food (RUTF) in Vietnam. *Food and nutrition bulletin*, *35*(2_suppl1), S52-S56.

Kevin Rennert, K., & Kingdon, C. (2019). Social Cost of Carbon 101. A review of the social cost of carbon, from a basic definition to the history of its use in policy analysis. *Resources for the Future,* https://www.rff.org/publications/explainers/social-cost-carbon-101/

Ritchie, H. & Roser, M. (2020). Environmental impacts of food production, https://ourworldindata.org/environmental-impacts-of-food' (accessed on 20, June, 2021).

Rohmer, S.U.K., Gerdessen, J.C. & Claassen, G.D.H. (2019). Sustainable supply chain design in the food system with dietary considerations: A multi-objective analysis. *European Journal of Operational Research*, *273*(3), 1149-1164.

Roth, A.V., Tsay, A.A., Pullman, M.E. & Gray, J.V. (2008). Unraveling the food supply chain: strategic insights from China and the 2007 recalls. *Journal of Supply Chain Management*, *44*(1), 22-39.

Santini, A., Novellino, E., Armini, V. & Ritieni, A. (2013). State of the art of Ready-to-Use Therapeutic Food: a tool for nutraceuticals addition to foodstuff. *Food chemistry*, *140*(4), 843-849.

Swaminathan, J.M. (2018). Big data analytics for rapid, impactful, sustained, and efficient (RISE) humanitarian operations. *Production and Operations Management*, *27*(9), 1696-1700.

So, A. & J.M. Swaminathan (2009). The Nutrition Articulation Project: A supply chain analysis of Ready to Use Therapeutic Foods to the Horn of Africa, UNICEF Technical Report.

Swaminathan, J. M., W. Gilland, V. Mani, C. M. Vickery, A. So. 2012. UNICEF employs prepositioning strategy to improve treatment of severely malnourished children. Working paper, Kenan-Flagler Business School, University of North Carolina, Chapel Hill

Tongwane, M., Mdlambuzi, T., Moeletsi, M., Tsubo, M., Mliswa, V. & Grootboom, L. (2016). Greenhouse gas emissions from different crop production and management practices in South Africa. *Environmental Development*, *19*, 23-35.

UNICEF(2020).https://data.unicef.org/country/ind/ ; https://www.financialexpress.com/opinion/un-report-on-under-5-mortality-india-one-of-the-worst-in-gender-disparities/1825685/ (accessed on 17 Jan, 2020).

UNICEF (2021a). Ready-to-use therapeutic food: current outlook. UNICEF Supply Division. March.

UNICEF (2021b). Pricing. UNICEF Supply Division. February.

Volpe, R., Messineo, S., Volpe, M. & Messineo, A. (2015). Carbon footprint of tree nuts-based consumer products. *Sustainability*, *7*(11), 14917-14934.

World Bank (1993). World development report 1993: investing in health. Washington, DC, accessed June 2, 2019, available at https://openknowledge.worldbank.org/handle/10986/5976.

World Bank (2018), https://www.worldbank.org/en/news/press-release/2018/10/17/nearly-half-the-world-lives-on-less-than-550-a-day.

World Food Programme (WFP). 2009a. Learning from Experience: Good Practices from 45 Years of School Feeding. WFP, Rome.

World Health Organization (2019). *Sustainable healthy diets: guiding principles*. Food & Agriculture

World Health Organization (WHO, 2021), https://www.who.int/news-room/fact-sheets/detail/malnutrition (accessed on 24, June, 2021)

Worldmeters (2021). https://www.worldometers.info/world-population/india-population/ (accessed on 24, June, 2021)

DECISION SCIENCES INSTITUTE

Does Underuse Variation in Test-Ordering Practice Relate to Higher Care-Delivery Cost?

Seokjun Youn
University of Arizona
Email: syoun@email.arizona.edu

Gregory R. Heim
Texas A&M University
Email: gheim@mays.tamu.edu

ABSTRACT

Recent studies estimate that a significant portion, approximately 25%, of total healthcare spending in the U.S. is waste, ranging from $690 billion to $935 billion every year. Naturally, there has been a sustained call for efforts to reduce such waste of healthcare resources. This study aims to empirically investigate how unwarranted clinical practice variation in healthcare relates to hospital resource usage. Specifically, we posit that hospitals with a higher underuse practice variation in test-ordering (e.g., radiology and laboratory tests) may face unexpected higher expenditures in subsequent care-delivery stages. We discuss our findings and policy implications based on a comprehensive inpatient data from the states of New York and Florida.

KEYWORDS: Healthcare; Test-Ordering Practice Variation; Care-Delivery Cost; Process Quality; Experiential Quality; Empirical Operations

DECISION SCIENCES INSTITUTE

Cloud Computing: Students' Intention to Learn

Robert Cutshall
Texas A&M University-Corpus Christi
Email: robert.cutshall@tamucc.edu

Chuleeporn Changchit
Texas A&M University-Corpus Christi
Email: chuleeporn.changchit@tamucc.edu

Charles Changchit
Texas A&M University-Corpus Christi
Email: charles.changchit@gmail.com

ABSTRACT

The explosive growth of data driven by information and communication has caused a need for improving the way the data are stored and retrieved. The demand to store and retrieve data effectively and efficiently has forced both companies and individuals to pay attention to cloud computing technology. Cloud computing technology has gained much attention in both commercial and academic settings. This study aims to examine how college students perceive cloud computing. In specific, the research questions focus on determining which factors drive their intention to learn the cloud computing concept. The study findings reveal several factors which academic institutes should consider before adding a cloud computing topic into their curriculum.

KEYWORDS: Cloud Computing, Technology Acceptance Model (TAM), Perceived Usefulness, Perceived Ease of Use, Perceived Security, Perceived Speed, Perceived Cost, Normative Beliefs, Technology Competency

INTRODUCTION

Cloud storage has seen an increasing rise in demand and diffusion (Burda & Teuteberg, 2016). The cloud is a unique platform for generating data and innovative solutions to leverage that data (Ikink, 2021). Cloud computing has profound impacts on today's business environment (Fan et al., 2015). This technology has attracted considerable attention because their services offer numerous benefits including its ability to provide faster on-demand infrastructure, self-service, and the independent ability to contribute to and access resources (Lee, 2019). It is increasingly growing in convenience as a ubiquitous network that requires little interaction between the cloud service provider and the user (Changchit & Chuchuen, 2018). This flexibility and independence can help save consumers time and stress by making the process simple and allowing the account design to be based entirely on individual needs.

The cloud computing model has become tremendously popular due to its benefits such as cost-effectiveness, scalability, usefulness, ease of use, and worldwide accessibility. Since 2010, the global cloud services industry has risen year-over-year to reach a $370 billion valuation in 2020, indicating a growth of over 380 percent in ten short years (Ikink, 2021). Cloud computing adoption is exploding, with enterprise applications migrating to the public cloud and organizations becoming more cloud-native in their deployments (Costello, 2021).

Cloud computing has attracted a significant amount of attention in both commercial and academic settings. Education is also a driving force for the continuous improvement of cloud computing (Changchit & Chuchuen, 2018). This technology provides a number of benefits to an educational setting (Behrend et al., 2011). Students can gain a lot from this technology as it serves as a convenient mobile storage space (Singh & Veralakshmi, 2012). Some of the benefits that students can derive from cloud computing, in an educational setting, include using it for completing assignments, online classes, group projects, creating and editing papers and presentations; as well as for work or entertainment (Changchit, 2015).

For students, the attractiveness of cloud computing is not only its cost effectiveness, but also the thirst for technology that college students have today, which makes learning and adopting these new technologies easier for them. Integrating cloud computing concepts and applications into the curriculum provides students with skill and competency development related directly to today's workplace (Mitchell & Meggisson, 2014). With the many benefits generated by cloud computing technology, it is interesting to find out how students perceive cloud computing and which factors can drive their intentions to learn this technology.

LITERATURE REVIEW

Over the past decade, the phrase “data is currency” has permeated corporate culture from Silicon Valley to Stockholm to Sydney (Ikink, 2021). Technological advances, such as mobile devices, social media, and cloud computing, help companies gain competitive and strategic advantages (Lee, 2019). The high demand for more advanced and efficient technology has contributed to the creation of new advancements in technology such as cloud computing. Cloud Computing has transformed the IT industry by opening the possibility for cloud computing services to deliver the enterprise applications and software through platforms such as Amazon Elastic Cloud, Microsoft's Azure, Google App Engine as a service (SaaS).

In academic settings, cloud computing is a promising prospect for educational institutions, especially during times of budget constraints. Prior research on cloud computing adoption in educational settings has focused most of its efforts on understanding the drivers and constraints that students and schools perceive in the adoption of this computing model. For example, Behrend et al (2011) examined the factors leading to adopting cloud computing as a virtual computing lab for a class. The finding revealed that students' ease of use perception would positively affect intentions for future use, but not for actual use. In addition, students who complete their work faster and in a more practical manner were more likely to recognize cloud computing as an effective service. This study also found that students with anxiety about new technologies had a negative effect on the perceived usefulness of new technologies. Another study also suggested that in order to deal with technology anxiety, it is important for universities to plan hands-on training to help students become more familiar with these new technologies (Blue & Tirotta, 2011).

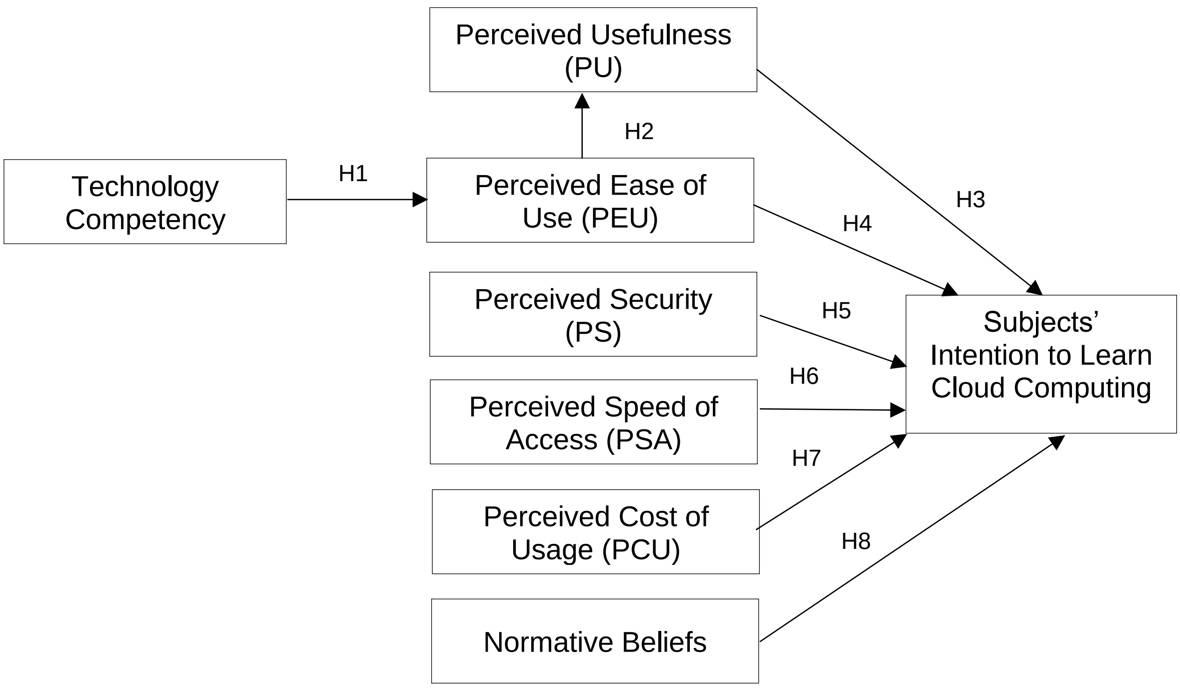
With many benefits promised by the use of cloud computing, it is not surprising that the demand for graduates with exposure to Cloud Computing is on the rise (Chen et al., 2012). Integrating cloud computing concepts and applications into the business curriculum provides skill and competency development related directly to today's workplace (Mitchell & Meggison, 2014). However, despite the many benefits of cloud computing, very few studies have focused on students' willingness to learn this technology. Changchit (2015) revealed factors that play an important role in encouraging students to accept cloud computing. These factors included perceived usefulness, perceived ease of use, perceived security, perceive speed of access, and perceived cost of usage. Another study investigated the beliefs and perceptions of school principals in Sweden toward cloud computing (Lim et al., 2015). Results suggested principals of Swedish schools believe the main benefits of cloud computing is its ability to allow users to access data and software anywhere and its ability to facilitate sharing of learning materials and data.

To increase the potential for success, it is crucial for any institution to understand which factors can motivate students to learn about cloud computing before integrating it into the curriculum. The findings in this study should help programs to focus on the factors that encourage students to develop an interest in the topic as well as finding ways to minimize the effect of discouraging factors, thus increasing the chance for the success of the integration of cloud computing into the program. Therefore, this study aims at examining the factors which impact students' willingness to learn cloud computing.

THEORETICAL DEVELOPMENT AND RESEARCH MODEL

The theoretical framework for this study is based on the Technology Acceptance Model (TAM). TAM was developed by Davis in 1986 to introduce the factors impacting users' acceptance of information technology **(**Davis et al., 1989). The model is widely used by many disciplines and has been extended to include more relevant factors relating to various research topics. This study modified the TAM to include five additional factors predicted to play an important role in motivating students' to learn cloud computing. These factors are: (1) Technology Competency (TC), (2) Perceived Security (PS), (3) Perceived Speed of Access (PSA), (4) Perceived Cost of Usage (PCU) and (5) Normative Beliefs. The proposed research model is shown in Figure 1 below.

Figure 1 Research Model



Technology Competency

Typically, people are more likely to use an information system if they have a higher level of technology competency (Yang, 2010). Technology Competency is defined as an individual's assessment of their capability to use computers in a variety of situations (Hsia et al., 2014). People with greater technology competency tend to have a larger motivation to use technology to accomplish tasks as they have a higher level of confidence in their ability to accomplish the tasks on a computer. In regard to learning about the cloud computing topic, it is likely that students who are more technology competent will perceive the technology as easy to use. We therefore propose that:

H1: Technology Competency positively affects subjects' perceived of ease of use.

Perceived Ease of Use of Cloud Computing (PEU)

Perceived ease of use plays a crucial role in understanding individual responses to information technology (Chau & Hu, 2001). This factor is defined as the degree to which the prospective user expects the target system to be free of effort (Davis et al., 1989). Based on the TAM, this factor has been shown to have an impact on perceived usefulness (Davis, 1986). In addition, research over the past decade provides evidence that this factor also has a significant effect on usage intention (Venkatesh & Davis, 2000). We thus hypothesize that:

H2: Perceived ease of use (PEU) positively affects subjects' perceived usefulness.

H3: Perceived ease of use (PEU) positively affects subjects' intention to learn cloud computing.

Perceived Usefulness of Cloud Computing (PU)

Davis et al (1989) defined perceived usefulness as the prospective users' subjective probability that using a specific application system will increase their job performance within an organizational context. This factor has a significant impact on usage intention (Davis et al., 1989; Venkatesh & Davis, 2000). A prior study also reported that this construct positively influenced subjects' intention to adopt cloud computing (Changchit & Chuchuen, 2018). Based on the foregoing, we propose the following hypothesis:

H4: Perceived usefulness (PU) positively affects subjects' intention to learn cloud computing.

Perceived Security of Cloud Computing (PS)

Flavián and Guinalíu, (2006) define perceived security as a subjective probability with which consumers believe that their personal information (private and monetary) will not be viewed, stored, and manipulated during transit and storage by inappropriate parties in a manner that is inconsistent with their expectations. Since security is always a major concern for all individuals who are dealing with sensitive data in everyday life (Changchit, 2008), there is a high likelihood that subjects would be more willing to learn about cloud computing if they find this technology to be secure. We therefore posit that:

H5: Perceived security (PS) positively affects subjects' intention to learn cloud computing.

Perceived Speed of Access of Cloud Computing (PSA)

For cloud computing to be widely accepted, it is crucial that the services must allow users to access data at a reasonable speed (Changchit & Chuchuen, 2018). The speed of using applications over the Internet can be a factor that has prevented cloud computing from being a viable option for outsourcing IT operations. Users may be unaware that the use of applications via the Internet still allow them to retrieve the data at the same speed as when the data is stored locally. Students' perception on the speed of access should influence their intention to learn about cloud computing. Hence, we posit that:

H6: Perceived speed of access (PSA) positively affects subjects' intention to learn cloud computing.

Perceived Cost of Usage of Cloud Computing (PCU)

In 2011, the monthly cloud expenditure of an average company was roughly $6,500. In 2020, this number had risen to $10,000 per month (Ikink, 2021). A survey conducted by ComputerWorld magazine (Wood, 2011) of IT professionals revealed that while "saves money" ranked first on the list of cloud computing key benefits, "costs more" ranked third on the list of drawbacks suggesting that the issue of cost in outsourcing information technology resources is a complex issue. In many public and private sector industries, including education, federal and state government, and telecommunications, cloud computing systems are being pilot-tested and implemented to save IT costs and improve performance (Behrend et al., 2011). Students perception of the cost of cloud computing should impact their willingness to learn this technology. We therefore hypothesize that:

H7: Perceived cost of usage (PCU) positively affects subjects' intention to learn cloud computing.

Normative Beliefs (NB)

Fishbein and Ajzen (1975) defined normative beliefs as individuals' perception that most people who are important to them think they should or should not perform the behavior in question. Limayem et al. (2000) have shown that normative beliefs are significant in the realm of online shopping. In regard to technology usage in an academic setting, Salleh (2016) found that normative beliefs have a positive influence on intention. Thus, we hypothesize:

H8: Normative beliefs (NB) positively affects subjects' intention to learn cloud computing.

RESEARCH METHODOLOGY

Development of Measurement Instrument

The measurement scales for this study utilized the TAM survey instrument and modified it with the addition of five constructs. The perceive ease of use and perceived usefulness factor survey items were revised from the work of Venkatesh & Davis (2000) and Venkatesh et al. (2003). Other survey items to measure the additional factors in this study were developed to identify the significant constructs leading to students' intention to learn about cloud computing. Several tests such as reliability, KMO and Bartlett's, common method bias, and factor analysis were conducted in this study to verify and validate their suitability for the measurement model in this study. These results are described in the data analysis section of this paper.

The questionnaire consists of forty-seven (47) questions. Forty questions with the five-point Likert scale were designed to measure subjects' perceptions on cloud computing and whether they believe it should be integrated into their curriculum. The remaining seven questions were asked to gather some demographic data on the subjects. To validate the clarity of these questions, three professors and three researchers were asked to read through the survey questions. Revisions to the survey were made based on the feedback received.

Data Collection

The surveys were administered to students at a Southern university in the United States These students are certainly part of the target group for companies providing cloud computing services. Four hundred and sixty-seven (467) subjects participated in this study. However, only four hundred and thirty-five (435) responses are valid. Details on the subjects' demographics are provided in Table 1 below.

Table 1: Subjects' demographics (n=435)

	No.	**%**		**No.**	**%**
Gender			**Ethnicity**		
Male	170	39.1	African	5	1.2
Female	223	51.3	Anglo	54	12.4
No answer	42	9.6	Asian	17	3.9
			Hispanic	92	21.2
Age			Native American	3	0.7
18 - 24	347	79.7	Other	0	0.0
25 - 34	48	11.0	No Answer	264	60.6
35 - 44	9	2.1			
45 and over	6	1.4	**Classification**		
No answer	25	5.8	Freshman	77	17.7
			Sophomore	136	31.3
First Generation College Student			Junior	115	26.4
Yes	158	36.3	Senior	41	9.4
No	217	49.9	Graduate	23	5.3
No answer	60	13.8	Other	0	0.0
			No answer	43	9.9
College					
Business	217	49.9	**Employment Status**		
Education	34	7.8	Full-time	45	10.3
Liberal Art	56	12.9	Part-Time	196	45.1
Nursing	38	8.7	Un-employed	151	34.7
Science & Technology	43	9.9	No Answer	43	9.9
No answer	47	10.8			

DATA ANALYSIS AND DISCUSSION

The statistical software package SPSS 26 with AMOS 25 was used to analyze the respondents' data. For SEM, Hair et al. (2009) recommends a sample size of at least 200. Therefore, our sample of 435 exceeds the recommended value.

Reliability and Multicollinearity Test

A reliability test was conducted to examine the internal consistency of the research instrument. The test confirmed the reliability with Cronbach's alpha coefficient of 0.961. In addition, since multicollinearity can have harmful effects (Cenfetelli & Bassellier 2009), multicollinearity was assessed for all the indicators in the research model. The results revealed that the multicollinearity is not an issue with this data set.

KMO and Bartlett's Test

As shown in Table 2 below, the KMO and Bartlett's Test was conducted to assess the degree of unidimensionality of the scales. The test confirmed the sampling adequacy with the value of 0.947. The Bartlett's test of sphericity showed a p-value of 0.000 for both sets of data. Thus, the null hypothesis was rejected regarding no difference between the correlation matrix and the identity matrix.

Table 2. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.947
Bartlett's Test of Sphericity	Approx. Chi-Square	11002.483
	df	561
	Sig.	0.000

Common Method Bias

To ensure that the model is free from common method bias, which is a measurement error that threatens the validity of conclusions drawn from statistical results, Harman's single factor test was conducted. The Harman's single factor test which is the most widely used in the literature (Roni, 2014) was conducted and was obtained by running an un-rotated, single-factor constraint of factor analysis in SPSS statistics. As shown in Table 3 below, the variance explained by a single factor of 41.498% is less than the recommended 50% cut-off point (Roni, 2014), indicating that the common method bias is not a major concern in this study.

Table 3. Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	14.109	41.498	41.498	14.109	41.498	41.498	4.248	12.493	12.493
2	2.932	8.624	50.122	2.932	8.624	50.122	3.949	11.613	24.107
3	2.030	5.971	56.093	2.030	5.971	56.093	3.687	10.844	34.951
4	1.610	4.736	60.830	1.610	4.736	60.830	3.445	10.133	45.084

5	1.478	4.347	65.176	1.478	4.347	65.176	2.940	8.648	53.732
6	1.291	3.798	68.974	1.291	3.798	68.974	2.770	8.147	61.879
7	1.258	3.699	72.674	1.258	3.699	72.674	2.426	7.135	69.014
8	1.011	2.973	75.647	1.011	2.973	75.647	2.255	6.633	75.647
9	.612	1.799	77.446						
10	.576	1.693	79.139						
11	.535	1.573	80.712						
12	.490	1.441	82.153						
13	.435	1.281	83.434						
14	.415	1.220	84.654						
15	.408	1.199	85.853						
16	.386	1.135	86.988						
17	.367	1.079	88.067						
18	.350	1.029	89.097						
19	.346	1.019	90.115						
20	.313	.921	91.036						
21	.294	.865	91.901						
22	.285	.839	92.741						
23	.271	.797	93.538						
24	.261	.767	94.305						
25	.246	.724	95.029						
26	.233	.685	95.714						
27	.221	.651	96.365						
28	.209	.616	96.981						
29	.202	.595	97.576						
30	.192	.566	98.142						
31	.187	.551	98.693						
32	.165	.485	99.178						
33	.149	.437	99.615						
34	.131	.385	100.000						

Extraction Method: Principal Component Analysis.

Factor Analysis

Confirmatory factor analysis with varimax rotation was conducted to examine the construct validity and to verify the groupings of the survey items adopted from previous studies. The results of the factor analysis confirm that the forty survey items distributed themselves into eight factors (see Table 4). The survey items which recorded a value below the suggested reliability level of 0.65 (Hair et al. 2009) were removed from the data analysis.

Table 4: Factor Analysis

Constructs	Component							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Perceive Usefulness (PU1)	.139	.296	.221	.262	.205	**.667**	.081	.145
Perceive Usefulness (PU2)	.125	.169	.122	.283	.083	**.774**	.132	.127
Perceive Usefulness (PU3)	.225	.286	.241	.163	.218	**.694**	.094	.034
Perceive Usefulness (PU4)	.199	.136	.176	.197	.220	**.681**	.051	.096
Perceived Ease of Use (PEU1)	.243	.214	.218	.110	**.724**	.240	.129	.098

Perceived Ease of Use (PEU2)	.288	.228	.218	.070	**.767**	.230	.132	.003
Perceived Ease of Use (PEU3)	.240	.164	.241	.119	**.767**	.171	.102	.017
Perceived Ease of Use (PEU4)	.147	.176	.097	.168	**.709**	.061	.177	.222
Perceived Security (PS1)	.121	.172	.211	.235	.190	.067	**.781**	.090
Perceived Security (PS2)	.155	.192	.167	.171	.138	.079	**.844**	.104
Perceived Security (PS3)	.164	.183	.275	.096	.127	.148	**.764**	.091
Perceived Speed of Access (PSA1)	.180	**.769**	.190	.178	.152	.193	.106	.115
Perceived Speed of Access (PSA2)	.203	**.775**	.288	.142	.132	.197	.098	-.002
Perceived Speed of Access (PSA3)	.189	**.773**	.160	.092	.201	.115	.192	.119
Perceived Speed of Access (PSA4)	.216	**.777**	.222	.148	.184	.199	.186	.098
Perceived Speed of Access (PSA5)	.195	**.719**	.261	.201	.170	.161	.133	.143
Perceived Cost of Usage (PCU1)	.086	.159	**.734**	.154	.138	.146	.184	.072
Perceived Cost of Usage (PCU2)	.170	.200	**.739**	.088	.163	.191	.162	.112
Perceived Cost of Usage (PCU3)	.148	.282	**.725**	.159	.180	.146	.164	.138
Perceived Cost of Usage (PCU4)	.150	.283	**.699**	.216	.132	.180	.122	.176
Perceived Cost of Usage (PCU5)	.204	.197	**.711**	.216	.204	.099	.174	.170
Normative Beliefs (NB1)	-.010	.211	.258	.179	.130	.081	-.005	**.735**
Normative Beliefs (NB2)	-.086	.038	.194	.184	.152	.057	.121	**.823**
Normative Beliefs (NB3)	-.032	.075	.027	.110	-.014	.131	.111	**.811**
Technology Competency (TC1)	**.811**	.214	.099	.237	.152	.109	.030	-.039
Technology Competency (TC2)	**.754**	.165	.130	.088	.187	.082	.171	-.097
Technology Competency (TC3)	**.845**	.171	.157	.077	.199	.113	.092	-.019
Technology Competency (TC4)	**.855**	.145	.152	.170	.125	.174	.086	-.003
Technology Competency (TC5)	**.843**	.146	.112	.165	.155	.146	.120	.037
Intention to Learn (INT1)	.138	.101	.034	**.755**	.090	.061	.135	.183
Intention to Learn (INT2)	.106	.187	.136	**.780**	.109	.105	.082	.098
Intention to Learn (INT3)	.197	.187	.251	**.699**	.044	.275	.025	.077
Intention to Learn (INT4)	.116	.028	.154	**.678**	.048	.252	.184	.116
Intention to Learn (INT5)	.213	.186	.223	**.700**	.194	.209	.133	.120

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

The loadings are <0.65 were removed from the data analysis.

Structure Equation Model

SPSS AMOS 25 was utilized to evaluate the proposed research model. To test the overall goodness of fit of the proposed research model, the measures of df/Chi-square, Goodness of fit, Adjusted goodness of fit, Root mean square error of approximation, Comparative fit index, Tucker Lewis index, and Normed fit index were employed. Table 5 reveals that all the goodness of fit indices fall within their acceptable levels (Bentler & Bonett, 1980; Hu & Bentler, 1999; Tucker & Lewis, 1973). This reveals that the proposed research model exhibited a good fit with the data.

Table 5. Fit Indices for the Models

Fit Indices	Recommended Value	Measurement Model
Chi-square (CMIN)/df	<=3.00	1.282
Goodness-of-fit (GFI)	>=0.90	0.999
Adjusted goodness-of-fit (AGFI)	>=0.80	0.999
Normed fit index (NFI)	>=0.90	0.997
Tucker Lewis Index (TLI)	>=0.90	0.995
Comparative fit index (CFI)	>=0.93	0.999
Root Mean Square Error of Approximation (RMSEA)	<=0.06	0.025

Hypothesis Testing

Properties of the causal paths including standardized path coefficients are presented in Figure 2. The results of hypothesis testing are shown in Table 6.

Figure 2 Structural equation model path analysis

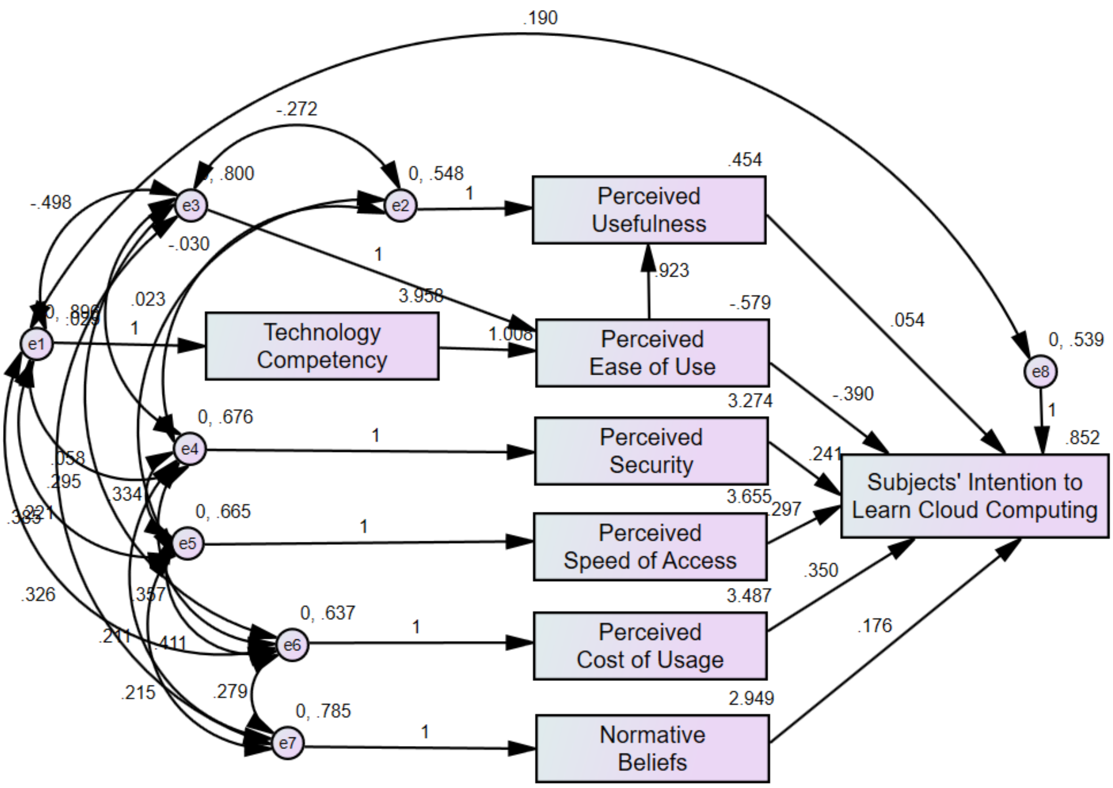


Table 6: Hypothesis testing and results

H#	Hypothesis Testing			Standardized estimate (β)	Critical Ratio	p-value
1	Technology Competency	→	Perceived Ease of Use of Cloud Computing	1.008	11.102	***
2	Perceived Ease of Use of Cloud Computing	→	Perceived Usefulness of Cloud Computing	0.923	14.371	***
3	Perceived Ease of Use of Cloud Computing	→	Intention to Learn Cloud Computing	-0.390	-2.795	0.005
4	Perceived Usefulness of Cloud Computing	→	Intention to Learn Cloud Computing	0.054	0.406	0.685
5	Perceived Security of Cloud Computing	→	Intention to Learn Cloud Computing	0.241	4.000	***
6	Perceived Speed of Access of Cloud Computing	→	Intention to Learn Cloud Computing	0.297	2.908	0.004
7	Perceived Cost of Usage of Cloud Computing	→	Intention to Learn Cloud Computing	0.350	3.643	***
8	Normative Beliefs	→	Intention to Learn Cloud Computing	0.176	4.015	***

*** indicates significance level < 0.001

For hypothesis H1, the result demonstrates a significant relationship between subjects' technology competency and their perceived ease of use of cloud computing (β = 1.008, p-value < 0.001). This finding indicates that subjects who consider themselves confident with using computers also find it easy to use cloud computing. The higher the perceived level of technology competency that a student has, the more likely that the student will find cloud computing easy to use (Venkatesh & Davis, 1996).

The result of hypothesis H2 confirms the relationship between perceived ease of use and perceived usefulness (β = 0.923, p-value < 0.001). The finding is in line with the study conducted by Davis et al. (1989), which indicates that the perceived ease of use can positively influence subjects' perceived usefulness.

Hypothesis H3 examines the relationship between perceived ease of use of cloud computing and students' intention to learn cloud computing. The result reveals a significant negative relationship between these two constructs (β = -0.390, p-value = 0.005). This result confirms the findings of some prior studies that found a significant relationship between perceived ease of use and intention (e.g., Dasgupta, et al. 2011; Sripalawat et al. 2011) but contradicts other previous studies that reported no significant relationship between perceived ease of use and intention (e.g., Koenig-Lewis et al. 2010; Liébana-Cabanillas et al. 2016). It is interesting that the data appear to suggest that the higher the perceived ease of use is the less likely students will have the intention to use cloud computing. Therefore, the students' intention to learn cloud computing will be lower if they believe that it is easy to use cloud computing.

Regarding hypothesis H4, it is quite interesting to find that the factor perceived usefulness of cloud computing does not have a significant influence on students' intention to learn cloud computing (β = 0.054, p-value = 0.685). This finding contradicts to earlier research (e.g., Hanafizadeh et al. 2014; Moqbel & Bartelt, 2015; Püschel et al. 2010). Although the finding is similar to the results of a study by Lucia-palacios, Pérez-lópez, & Polo-redondo, (2016) who reported that perceived usefulness is not a significant factor in the adoption of cloud computing in a model that includes inertia. It appears that whether students believe that cloud computing is useful or not, it does not impact their intention to learn cloud computing. This could be due to the fact that students may know that businesses are increasingly turning to the cloud to support their IT function. Thus, it is irrelevant what the students think as they know they will have to learn to use cloud computing if they go to work of a business that employees cloud computing.

Hypothesis H5 reveals the relationship between the perceived security of cloud computing and students' intention to learn cloud computing (β = 0.241, p-value < 0.001). This result validates prior studies that perceived security plays an important role in technology acceptance (Alalwan et al. 2017; Bhatt. 2016; Svilar & Zupančič. 2016). It is apparent that students are willing to learn cloud computing if they believe this technology is secure.

For hypothesis H6, the results confirm that there is a significant relationship between the perceived speed of access of cloud computing and students' intention to learn cloud computing (β = 0.923, p-value = 0.004). It is quite obvious that the speed of access should be considered a crucial feature of cloud computing since slow access would delay students' completion of their work. Subjects who view cloud computing as having good speed of access are more likely to want to learn about this technology and thus willing to accept it as part of their curriculum.

The results demonstrate support for H7. The perceived cost of usage of cloud computing is also positively related to students' intention to learn cloud computing (β = 0.350, p-value < 0.001). This finding indicates that subjects who perceive the cost of cloud computing as low are willing to learn more about the cloud computing technology and are likely to accept it as part of their core curriculum.

Lastly, hypothesis H8 reports a positive relationship between normative beliefs and students' intention to learn cloud computing (β = 0.176, p-value < 0.001). This finding conforms to the results in a prior study by Bock et al.(2005). Since students usually need to share data among friends and family members, they are more likely to be heavily influenced by the experiences of these people.

CONCLUSION AND STUDY IMPLICATIONS

The results in this study reveal that six out of the seven factors play an important role in influencing students' intention to learn cloud computing either through an indirect or direct relationship. The results of the data analysis reveal support for hypotheses H1-H3 and H5-H8 (see Table 6). Hypothesis H4 was not supported. Overall, the independent variables in the research model explain 75.647 percent of the variability in subjects' intention to learn cloud computing.

The factors which support results are technology competency, perceived ease of use, perceive speed of access, perceived cost of usage, and normative beliefs. Interesting, this study does not find a significant relationship between perceived usefulness and students' intention to learn cloud computing. This result reveals that regardless of whether students perceive cloud computing as a technology that will be beneficial to them, they will be generally open to learning this cloud computing technology since businesses are increasingly adopting cloud computing.

The results in this study reveal that students with higher competence in technology will find cloud computing easier to use than those who have lower technology competency and thus influence their intention to learn this technology. In addition, if students believe that the cloud computing is secure, they will be more like to want to learn this technology.

Findings in this study also demonstrate that the speed of access and the cost of using cloud storage are also crucial factors that can influence students' intention to learn cloud computing. Students now prefer technology that allows them to get the thing they want in a timely manner. In addition, it does not matter how great a service a new technology can render, it will be useless if students cannot afford the cost of using it. Students will be more likely to accept cloud computing technology if they perceive that it does not cost them much to use the technology.

Based on the findings, normative beliefs also positively influence students' intentions to learn cloud computing. This is quite understandable as students will need to share data with their family members and friends. The attitudes of people around them should influence their intention to learn the cloud computing technology. Interestingly, this study does not find a relationship between students' willingness to learn cloud computing and their perception about its usefulness. This could be due to the fact that students may know that businesses are increasingly using the cloud to store data. It is therefore irrelevant what they think as they know they will have to learn to use cloud computing as required by the industries.

As in most empirical studies, there is an inherent limitation in this research. The sample in this research was limited to subjects from one university. Although there was an attempt to gather the data from a variety of courses in the university, future research should be conducted at multiple universities. Further research should also consider expanding demographics to include students in various countries. In addition, a future study could investigate in more detail which feature of cloud computing help increase students' effectiveness and efficiency in the classroom.

REFERENCES

Alalwan, A. A., Dwivedi, Y.K., & Rana, N.P. (2017). Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust. *International Journal of Information Management*, 37(3), 99-110.

Behrend, T., Wiebe, E.N., London, J., & Johnson, E. (2011). Cloud computing adoption and usage in community colleges*. Behavior & Information Technology*, 30(2), 231-240.

Bentler, P., & Bonett, D. (1980). Significance tests and goodness-of-fit in analysis of covariance structures. Psychological Bulletin, 88(3), 588-606. DOI:10.1037/0033-2909.88.3.588.

Bhatt, A. (2016). Factors affecting customer's adoption of mobile banking services. *Journal of Internet Banking and Commerce*, 21(1), 1-22.

Bock G.W., Zmud, R.W., Kim Y.G., & Lee J.N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *MIS Quarterly*, 29(1), 87–111.

Burda, D. & Teuteberg, F. (2016). Exploring consumer preferences in cloud archiving – a student's perspective. *Behaviour & Information Technology*, 35(2), 89-105.

Cenfetelli, R. T., & G. Bassellier. G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly,* 33(4), 689–707.

Changchit, C. (2008). Data protection and privacy issue. *Journal of Information Privacy and Security*, 9(3), 1-2.

Changchit C. & Chuchuen, C. (2018). Cloud computing: An examination of factors impacting users' adoption, *Journal of Computer Information Systems*, 58(1), pp. 1-9.

Chau, P.Y.K., & Hu, P.J.H. (2001). Information technology acceptance by individual professionals: a model comparison approach. *Decision Sciences*, 32, 699-719.

Chen, L., Liu, Y., Gallagher, M., Pailthorpe, B., Sadiq, S., Shen, H. T., Li, X. (2012). Introducing cloud computing topics in curricula. *Journal of Information Systems Education*, 23(3), 315-324.

Costello, K. (2021). Gartner Predicts the Future of Cloud and Edge Infrastructure. Retrieved from https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-cloud-and-edge-infrastructure/

Dasgupta, S., Paul, R., & Fuloria, S. (2011). Factors affecting behavioral intentions towards mobile banking usage: Empirical evidence from India. *Romanian Journal of Marketing*, 6(1), 6-28.

Davis, F.D. (1986). A technology acceptance model for empirically testing new end-user information systems: theory and results. Unpublished Doctoral dissertation, MIT Sloan School of Management, Cambridge, MA.

Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management Science*, 35, 982-1003.

Fan, Y., Wu, C., Chen, C., & Fang, Y. (2015). The effect of status quo bias on cloud system adoption. *The Journal of Computer Information Systems*, 55(3), 55-64.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley, Reading, MA, USA.

Flavián, C., & Guinalíu, M. (2006). Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. *Industrial Management & Data Systems*, 106, 601-620.

Hair, J., W. Black, B. Babin, & R. Anderson. 2009. *Multivariate data analysis: A global perspective*. 7th ed. Upper Saddle River: Prentice Hall.

Hanafizadeh, P., Keating, B.W., & Khedmatgozar, H.R. (2014). A systematic review of Internet banking adoption. *Telematics and Informatics*, 31(3), 492–510.

Hsia, J.W., Chang, C.C. & Tseng, A.H. (2014). Effects of individuals' locus of control and computer self-efficacy on their e-learning acceptance in high-tech companies. *Behaviour & Information Technology*, 33(1), 51-64

Hu, L.T., & Bentler, P. M. (1995). Evaluating model fit. In Hoyle R.H. (Ed.), Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage.

Ikink, R. (2021). 25 cloud trends for 2021 and beyond. Retrieved from https://www.accenture.com/nl-en/blogs/insights/cloud-trends

Koenig-Lewis, N., Palmer, A., & Moll, A. (2010). Predicting young consumers' take up of mobile banking services. *International Journal of Bank Marketing*, 28(5), 410-432.

Lee, Y. (2019). Adoption intention of cloud computing at the firm level. *Journal of Computer Information Systems*, 59(1), 61-72, DOI: 10.1080/08874417.2017.1295792

Liébana-Cabanillas, F., Alonso-Dos-Santos, M., Soto-Fuentes, Y., & Valderrama-Palma, V.A. (2016). Unobserved heterogeneity and the importance of customer loyalty in mobile banking. *Technology Analysis & Strategic Management*. DOI: 10.1080/09537325.2016.1262021.

Lim N., Grönlundb, A., & Andersson, A. (2015). Cloud computing: The beliefs and perceptions of Swedish school principals. *Computers & Education*, 84, 90-100.

Limayem, M. Khalifa, M., & Frini, A. (2000). What makes consumers buy from Internet? A longitudinal study of on-line shopping, *IEEE Transaction System Man, and Cybernetics*, 30(4), 421–432.

Lucia-palacios, L., Pérez-lópez, R., & Polo-redondo, Y. (2016). Enemies of cloud services usage: Inertia and switching costs. *Service Business*, 10(2), 447-467. doi:http://dx.doi.org/10.1007/s11628-015-0277-y

Mitchell, R. & Meggison, P. (2014). Strategies for Integrating Cloud Computing Concepts. *Journal of Applied Research for Business Instruction*, 12(2), 1-6.

Moqbel, M.A. & Bartelt, V.L. (2015). Consumer acceptance of personal cloud: Integrating trust and risk with the technology acceptance model. *AIS Transactions on Replication Research*, 1, 5. doi:http://dx.doi.org/10.17705/1atrr.00005

Püschel, J., Mazzon, J.A., & Hernandez, J.M.C. (2010). Mobile banking: Proposition of an integrated adoption intention framework. *International Journal of Bank Marketing*, 28(5), 389-409.

Roni S.M. (2014). Introduction to SPSS. Researchgate. 2014 June. Retrieve from http://www.researchgate.net/publication/262151892_Introduction_to_SPSS.

Ryan, M.J. & Bonfield, E.H. (1980). Fishbein's intentions model: a test of external and pragmatic validity. *Journal of Marketing*, 44, 82-95.

Salleh, S. (2016). Examining the influence of teachers ' beliefs towards technology integration in classroom. *The International Journal of Information and Learning Technology*, 33(1), 17-35.

Singh, S.P, & Veralakshmi, R.S.R. (2012). Cloud computing: A promising economic model for library and information centers. *DESIDOC Journal OF Library & Information Technology*, 32, 526-532.

Sripalawat, J., Thongmak, M., & Ngramyarn, A. (2011). M-banking in metropolitan Bangkok and a comparison with other countries. *Journal of Computer Information System*, 51(3), 67-76.

Svilar, A., & Zupančič, J. (2016). User experience with security elements in Internet and mobile banking. *Organizacija*, 49(4), 251-260.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1-10. https://doi.org/10.1007/BF02291170

Venkatesh, V., & Davis, F.D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*. 27(3), 425-478.

Venkatesh, V., & Davis, F.D. (2000). “A theoretical extension of the technology acceptance model: four longitudinal field studies. *Management Science*, 46(2), 186-204.

Venkatesh, V., Morris, M.G., Davis, G.B. & Davis, F.D. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27, 425-478.

Wood, L. (2011). Cloud storage a steep climb. *ComputerWorld*, 20-30.

Yang, K. (2010). The effects of technology self-efficacy and innovativeness on consumer mobile data service adoption between American and Korean consumers. *Journal of International Consumer Marketing*, 22(2), 117–127.

DECISION SCIENCES INSTITUTE

A Comparison of the Traversal and Return Routing Policies for Lift Truck Order Picking

Vitaly Brazhkin
University of West Florida
Email: vbrazhkin@uwf.edu

ABSTRACT

This warehouse operations study addresses order picking routing optimization in a wide aisle. For the first time, a realistic machine movement is simulated for the two most common heuristic routing policies: return and traversal. Unlike previous research that favored the traversal policy for on-foot picking, this study shows that the traversal routing policy is superior for low pick densities of up to 12%, while the return policy is more efficient for higher pick densities. This work gives industrial engineers and warehouse managers a critical piece of information for planning warehouse layouts and optimizing picking operations in a modern mechanized warehouse.

KEYWORDS: Warehouse operations, Order picking, Routing policy, Lift truck, Simulation

INTRODUCTION

In 1988, Goetschalckx and Ratliff published their influential paper “Order Picking in an Aisle” where they compared the most popular order picking policies in a typical wide-aisle warehouse with completely manual order picking. A lot of things have changed since then in warehousing, and some have not received proper coverage in academic research (de Koster et al, 2007). So it is a good time to revisit the topic. This paper compares the two most popular policies identified by Goetschalcks & Ratliff (1988b) again, given the current state of warehousing operations, specifically for order picking performed with a lift truck. Additionally, we take an applied perspective and aim to provide clear managerial guidance when it comes to the choice of a routing policy in a modern mechanized wide-aisle warehouse.

Since 1980-s, logistics and warehousing have experienced dramatic changes. They have become more powerful and moved from being peripheral activities to being increasingly central and important for the economy (Iyengar et al, 2012). Both managers and academics started to view warehouses not as merely cost centers but as a part of a broader business strategy of an organization (Murphy & Poist, 1992; McGinnis & Kohn, 1988; McGinnis et al, 1987). Warehouses have become a part of the value creation chain and contribute to the cost leadership and service differentiation strategies identified by Porter (1985) through advantages in operating costs and improved service quality (Richards, 2017; Stank et al, 1994; Murphy & Poist, 1992; McGinnis & Kohn, 1988).

Over the past three decades, many things have changed in the business environment of warehouses. Rising customer expectations and costs of money and labor, fast development of technology and IT systems, global competition, and shifts of power and new distribution channels in supply chains led to a dramatic increase in productivity, level of mechanization and automation, reliance on information technology systems, expanded menus of services, improved service quality, and reduction of lead times and order processing costs (Brazhkin, 2018a; Richards, 2017; Faber et al, 2002; Stank et al, 1994; Raney & Walter, 1992; Dadzie & Johnston, 1991).

However, warehouses are under constant pressure to deliver more with less, or more specifically, to process and ship orders faster at a lesser cost (Guia & Tan, Jr., 2019; Newsome et al, 2013; Frazell, 2002; Peschke, 2001). One of the most significant opportunities to achieve this goal is order picking optimization (Guia & Tan, Jr., 2019). Order picking is a process of retrieving items from storage locations to meet a specific demand (Frazelle, 2002). In a typical warehouse, picking may account for a half of operating costs or more, so efforts of warehouse engineers and managers have long been focused on the efficiency of this activity (Frazelle, 2002). Order picking is highly labor intensive and time consuming, with one component, travel time to, between and from storage locations, accounting for 50 percent of the order picking time (Tompkins et al, 2003). Given its share in costs and lack of adding customer value, travel time is a natural first candidate for improvement (de Koster et al, 2007).

Academic research on finding more efficient ways of order picking, including travel time reduction, has been published since 1970-s. A typical convention for finding the optimal solution for the shortest travel time has been to find the shortest travel distance for an order picker (de Koster et al, 2007). Then the problem resembles the well-known traveling salesman problem (de Koster et al, 2007). However, it is rather hard to solve optimally for all situations and many heuristic solutions to minimize travel distance for specific warehouse settings were proposed over years (Masae et al, 2020). The heuristic solutions may be less efficient than the optimal solutions but they are easier to understand and implement (Masae et al, 2020).

Most recently the academic interest to the problem has spiked. According to Masae et al (2020), who published a systematic literature review on the subject, the number of peer-reviewed journal publications that proposed a *new* way of efficiently routing the order picker in a warehouse, i.e. a new routing policy, went up from 20 in the period of 30 years (1983-2012) to 34 in just 2013-2019. The number of papers that advanced a *previously proposed* routing policy doubled from 41 in 2007-2012 to 82 in 2013-2019 (Masae et al, 2020). It has been suggested that the increased academic interest to the problem reflects the mounting real world cost, time and customer service pressures that warehouses experience today (Masae et al, 2020; Newsome et al, 2013). Thus, an efficient routing policy for order picking is a current topic for both practitioners and academics that has only increased in importance in the time since the famous Goetschalckx and Ratliff (1988b) paper was published.

LITERATURE REVIEW AND PAPER POSITIONING

With hundreds of papers published on the subject, for a literature review of individual papers, the interested reader is referred to the order picking optimization literature reviews of de Koster et al (2007) and Masae et al (2020). Here we provide a summary of the relevant features of the routing policy optimization papers as a group to explain the positioning of our paper in this research stream.

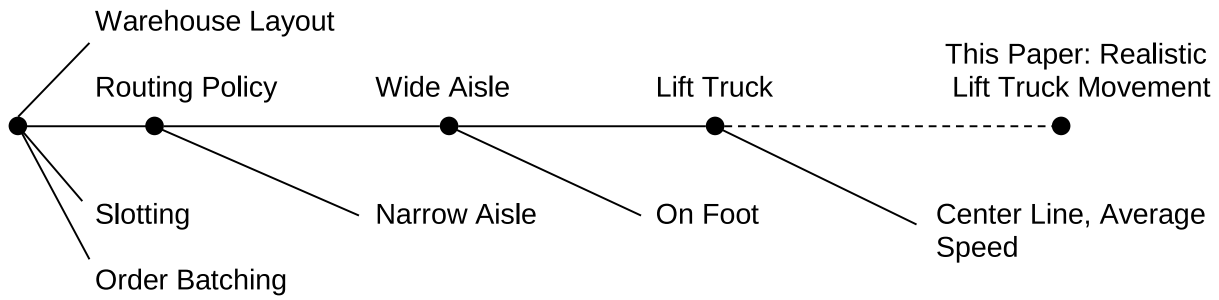
First, an important background knowledge fact: the type of a warehouse from the handling perspective matters. Handling of items in a warehouse can be manual, mechanized or automated. For example, many large retail chain distribution centers use miles of conveyor lines to sort fast moving consumer packaged goods. Automated systems are used to put away, store and retrieve most cargoes, except very small or oversized, in large airline cargo handling centers. Automation allows achieving higher productivity but comes with a high cost (Faber et al, 2002; Frazelle, 2002). There are also whole industries, such as furniture or automotive, that deal with slow moving, oddly-shaped nonconveyable items, for which automation is not an option (Brazhkin, 2018b). Some warehouses use manual or mechanized order picking for reasons of flexibility or lack of economies of scale to justify a high-cost investment in automation (Grosse et al, 2015).

Most warehouses use mechanized systems (Bowersox et al, 2013). In the order picking process, it means that human labor is combined with the use of handling equipment, such as lift trucks. However, there appears to be a disconnect between the actual warehousing handling practices and research. Previous research has overwhelmingly considered routing policy choices for an order picker traveling on foot (Masae et al, 2020). The lack of attention to the use of lift trucks in research has been noted (Burinskiene, 2011, 2015). Most papers that did consider order picking with a lift truck assumed narrow aisles where the operator can pick items on both sides of the aisle without maneuvering or leaving the machine (Masae et al, 2020).

This leaves relatively few papers that considered order picking on a lift truck in a wide aisle, including another paper by Goetschalckx and Ratliff (1988a). These papers were based on the assumption that the lift truck travels along the center line of the aisle with stops for the operator to get off the machine and manually retrieve items from the storage locations on both sides of the aisle. It was also common to assume a constant speed of the lift truck when in motion. Research on order picking on a lift truck in a wide aisle without these limiting assumptions to fill this gap is identified as one of the future research opportunities in the order picking routing literature review by Masae et al (2020).

Thus, the current work appears to be the first to consider order picking performed entirely on a lift truck in a wide aisle, without the aforementioned limiting assumptions and without any on foot travel component. Figure 1 illustrates the contribution of this paper to the picking optimization literature stream.

Figure 1. Paper contribution to the picking optimization literature stream

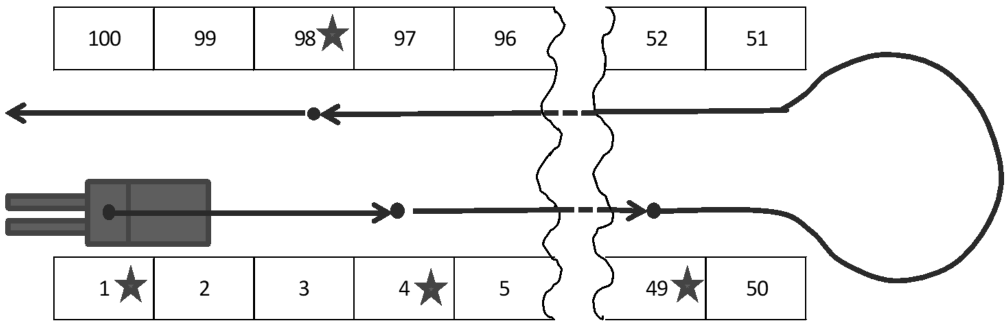


PROBLEM FORMULATION

The use of a lift truck introduces two main differences compared to the manual order picking: speed and driving path characteristics. A machine is clearly faster than a human. On the other hand, a human is more flexible. For example, a person can retrieve an item on one side of the aisle, turn around on the spot and then cross the aisle at the direct angle to pick an item directly across on the other side. A machine cannot do that. The lift truck driving path will be complex and will necessarily involve some backward movement thus increasing the route length and travel time. Thus, routing policies that favor aisle crossings will likely carry a time penalty associated with the complex machine travel path for this maneuver and may be less efficient than alternatives.

In traditional on foot routing optimization scenarios, the objective of finding the shortest route implies that the order picker, whether completely on foot or hopping on and off a lift truck, always moves with the same speed and there is no change of speed due to acceleration, deceleration or maneuvering. This is not the case for order picking performed entirely with a machine. Relevant lift truck speed characteristics include acceleration, deceleration and speed choices and limits. For example, maneuvers like turning around or crossing the aisle have to be performed at lower speeds for safety. On the other hand, routing policies that include longer stretches of linear forward motion may look more attractive as the lift truck can accelerate to its maximum speed and maintain it. Thus, modeling a realistic machine movement may change how routing policies compare among themselves.

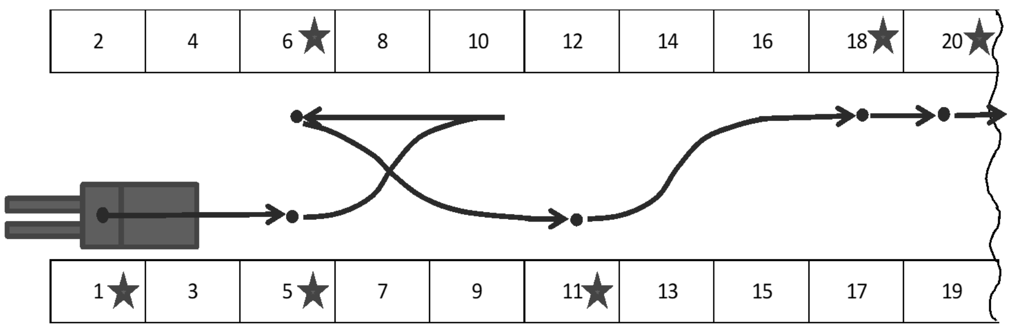
Figure 2. Example of the return routing policy (locations to visit are marked with stars)



The two most common routing policies are return picking and traversal picking (Goetschalckx & Ratliff, 1988b). The return policy requires that the order picker enter the aisle and visit picking locations on one side of the aisle first, then turn around and visit picking locations while returning to the entry point along the other side of the aisle. The picker never crosses the aisle except at the turnaround point. The return policy is graphically represented in Figure 2.

The traversal policy holds that the order picker enters the aisle, progressively visits locations on both sides of the aisle and then exits at the other end of the aisle rather than at the point of entry. The picker crosses the aisle as many times as necessary but never walks back. The lift truck version of the traversal policy with some unavoidable backward movement is graphically represented in Figure 3.

Figure 3. Example of the traversal routing policy (locations to visit are marked with stars)



Neither policy is always optimal. Their efficiency depends on the density of picks (number of picking locations the picker has to visit in an aisle). The traversal policy has an inherent advantage over the return policy in having the lift truck go through the aisle only once. It is particularly obvious if the scenario involves making only one pick (visiting only one location) in the aisle. However, the flip side of the coin is the need to cross the aisle to make multiple picks on both sides for typical picking scenarios. As the pick density increases, the time penalty for this complex maneuver performed by the lift truck multiple times will eventually outweigh the one time aisle passing advantage and become a disadvantage. This conjecture agrees with the findings by Goetschalcks & Ratliff (1988b) for picking on foot.

The work at hand models a realistic lift truck movement for both policies for different pick densities and compares the total time to complete the picking tour under either policy. The complexity of the maneuvers and the high variety of picking locations distribution scenarios do not permit a reasonable analytical estimation of the optimal scenario but the problem yields itself to finding heuristic solutions by other means. The method we used is explained next.

METHOD

Method Justification

The method used in this work is simulation. In early research, including warehouse operations research, simulation was not considered an independent method. It was primarily used to validate a proposed optimal solution determined through traditional mathematical and analytical means (Lee et al, 2013). Order picking routing optimization research routinely relied on many restrictive assumptions and simplifications (Kunder & Gudehus, 1975). Over time, the complexity of problems considered has grown, and so have the computational possibilities offered by advances in computer hardware and simulation software (Kelton, 2016; Kelton et al, 2015). However, analytical modeling research continued to remain narrowly focused on simplified optimal solutions and started to receive criticism for its lack of applicability to many real world warehouse scenarios (de Koster et al, 2007; Rowenhorst et al, 2000). Simulation, on the other hand, became an attractive method in itself. Its true appeal lies in its ability to tolerate realistic and complicated modeling assumptions to create a model of the complexity that cannot be approached by mathematical-analytical means, thus making simulation models more valid and accurate and ultimately more useful (Kelton, 2016). This led to a wider acceptance of simulation as a stand-alone research method in many areas, including supply chain management (Lee et al, 2013).

As noted in the Introduction, the picking route optimization problem is particularly complex, so the move away from optimal algorithms is quite evident. In a review of 203 academic papers on the matter, Masae et al (2020) found that 77% proposed a heuristic algorithm. In this paper, we compare the two most commonly used heuristic route policies while modeling a very true-to-life lift truck movement. Thus, our use of simulation as a method is justified and lies within the greater trend of the current warehousing research methodology.

Simulation Parameters

In the current and following sections we explain the simulation model parameters and simulation administration. We follow the methodological recommendations for computer simulations for SCM applications by Kelton (2016).

In general, the settings for the parameters were chosen to satisfy four conditions in the order of precedence: (a) be suitable for the simulation, (b) either coincide with or allow easy comparison to those used by Goetschalckx and Ratliff (1988b); (c) reflect real world scenarios and (d) be simple and easy to operate with and interpret. Two sets of parameters were used: one pertaining to the warehouse and the other—to the lift truck. They are given in Table 1 and explained in detail further.

A principle characteristic of a warehouse order picking system is an aisle width. This work, just as the paper by Goetschalckx and Ratliff (1988b), is one of the few that deals with order picking in a wide aisle. Goetschalckx and Ratliff (1988b) defined the aisle in relative units of slot widths, which is unacceptable for our simulation based on actual units of distance and time. Following other research of lift truck movement in a wide aisle, e.g., Burinskiene (2011, 2015) and Merkuryev et al (2009), we set the nominal aisle width at 5 meters. As we are not concerned with multiple aisles, the aisle width does not include the racks. We then calculate the effective width of the aisle to reflect that the movement path of the lift truck is defined by the movement of its geometrical center that never touches the sides of the aisle (the racks). Thus, the half-width of the order picking truck and a minimum safety gap on both sides need to be subtracted from the nominal aisle width. The narrowest order picking trucks are made to transport a Euro-pallet (1.2 m by 0.8 m) and correspondingly measure 0.8 m in width. The smallest routine operation safety gap between the stopped machine and the racks face was assumed to be 0.1 m, validated by actual picking operation observations. The resulting effective aisle width is then exactly 4 meters.

Table 1. Warehouse and lift truck data

PARAMETER	VALUE	MEASUR. UNIT
WAREHOUSE		
Effective aisle width	4	m
Aisle length	50	m
Total picking locations	100	
Distance between centers of two adjacent picking locations	1	m
Distance from entry or exit point to center of nearest location	1	m
LIFT TRUCK		
Maximum linear forward speed	2.8	m/s
Maximum speed for maneuvering and backing	1.4	m/s
Turnaround time, outside the aisle only	10	s
Acceleration	0.4	m/s^2
Deceleration	-0.4	m/s^2

The locations in the racks (slots) were chosen to be 1 m wide. The “clean” 4 to 1 ratio of the aisle and slot widths allows for a direct comparison with a scenario considered by Goetschalcks and Ratliff (1988b) and reflects some real world warehouse scenarios. The total number of locations in an aisle was set at 100 following prior research of Merkuryev et al (2009) and providing a convenient and realistic number to work with. Fifty 1-m slots on either side of the aisle determine the 50-m length of the aisle. The points of the aisle entry and exit for the lift truck were set to be 1 m away from the center of the first (or last) location. This completed the definition of the warehouse parameters for the simulation model.

Most lift truck parameters were chosen from the manufacturer's technical specifications and validated through an interview with a representative of a major lift truck manufacturer. While the parameters of a particular order picking truck make and model were chosen, the specifications for order pickers of other manufacturers are very close to the one chosen and to one another. According to the expert interviewed, this is in part dictated by human safety and comfort considerations. For example, given that the operator is standing, the order picking truck is engineered to accelerate smoothly, below its potential acceleration capability.

The speed parameters also agree with prior lift truck research, e.g., Burinskiene (2015). Additionally, order picking trucks are not allowed to turn around (make a U-turn) inside the aisle (for safety and lack of space reasons) and are required to exit the aisle, turn around and return. A typical real world scenario is used when the lift truck is required to come to a complete stop at the end of the aisle to determine that there is no cross-traffic before proceeding with the turnaround maneuver. Çelik and Süral (2016), who studied time penalties due to lift truck turns, reported that the speed may decrease by as much as 80 percent. Additional two seconds are imputed for the operator to establish the absence of cross traffic, but no actual cross-traffic or associated time penalty for yielding to it is assumed.

The single parameter calculated by the simulation was the total travel time on the picking tour in the aisle from the point of entry to the point of exit. While travel time is ultimately the metric that warehouses want to improve, most research, including Goetschalckx and Ratliff (1988b), used travel distance as a proxy for simplicity. This was possible because they assumed constant or average speed of the lift truck. We, on the contrary, model a realistic lift truck movement in the aisle and cannot replace time with distance. However, the comparison of the policies can still be made, and travel time optimization is the ultimate goal anyway.

A number of assumptions common for picking tour optimization studies were used in line with the work of Goetschalckx and Ratliff (1988b): (1) uncapacitated lift truck storage device (no premature returns because the container is full); (2) unconstrained pick sequence (the nature of items (e.g., heavy, fragile) to be picked did not affect the picking route), and (3) only travel time between the locations was considered (not the time to make an actual pick, which would depend on the item characteristics and quantity but not on the travel route).

Lift Truck Movement Modeling

A key difference of this work from previous research is modeling realistic lift truck movement. Both the movement path and speeds were taken into account.

The path of a lift truck inside the aisle was broken down into three principal components: linear forward movement, aisle crossing, and linear backward movement. (The U-turn outside the aisle for the return policy was described previously.) The travel time over each component was estimated separately and then added together to get the total travel time. A list of all possible scenarios of the combinations of movement directions and speeds is given in the Appendix.

Certain minor simplifications were made. In order to better reach items to be picked, operators position the lift truck very close to the face of the racks, with the gap as narrow as 0.1 m. When the lift truck is in motion, they move a little away from the racks for safety. Burinskiene (2011, 2015) used 0.3 – 0.6 m as a safety gap for a moving lift truck. Computational comparison of the travel times using the path that takes the lift truck up to 0.6 m away to the path continuing with the 0.1 m gap revealed a difference that was several orders below the rounding used in the total time estimation. Moreover, since it is not the absolute time that we were after but the comparison of the times for the two policies, and the gap would apply to both policies, we assumed that the truck always maintained the 0.1 m gap when in the stopped position and when moving along the racks in the aisle.

The path to cross the aisle was always modeled in a way that brought the lift truck to the closest possible point across from the starting position, given the turning radius identified in the lift truck specifications. This point, however, was never directly across from the starting position, so moving backwards was modeled to reach the positions that were skipped because crossing the aisle took the lift truck "too far." A sample movement path across the aisle is shown in Figure 3.

Two "cruising" speeds were modeled. A typical lift truck speed limit in a warehouse is 10 km/h (2.8 m/s) (Burinskiene, 2015), which, not coincidentally, was the maximum speed in the specifications for the order picking truck we used as well as many others we reviewed. Some lift truck models may be equipped with programs that allow using only reduced speeds for novice lift truck operators, depending on the code ID the operator enters to activate the machine. We assumed that all driving was performed by qualified experienced operators at maximum safe speeds. Furthermore, we excluded the effect of deliberate slowdown by operators due to pegging one's productivity to a perceived fair work standard, a practice described by Gray (1992), or any other reasons. If the distance between the positions allowed, the lift truck was estimated to reach the maximum speed after acceleration and maintained that speed until the point of starting deceleration needed to stop in front of the next position to visit on the picking tour. The maximum speed the lift truck was allowed to accelerate to when crossing the aisle or moving backwards was 50 percent of the maximum forward linear speed. If the positions were too close to each other to reach the maximum allowed speed, the travel distance was split in half between acceleration and deceleration.

Simulation Administration

Just as in the Goetschalckx and Ratliff (1988b) paper, the simulation was run for different pick density scenarios (share of picking locations to visit in an aisle on a picking tour). We started with a pick density of 5 percent and increased the pick density in 5 percentage point increments. Once the ballpark region of the two policies producing similar results became known, additional simulations were run to determine the closest pick density value to the point of indifference between the two routing policies we compared. Running simulations beyond the pick density of 25 percent was deemed unnecessary because very high pick densities are more likely to use an entirely different picking method, e.g., zone picking, and would not have added new information to the evaluation of the two routing policies.

The specific picking locations to visit for a given pick density were randomly generated using the RAND function in Miscrosoft Excel. There were concerns about a potential bias of the random number generator in earlier versions of Excel prior to 2010 but not in later versions (Kelton, 2016). We used the 2019 edition of Excel for this simulation.

Kelton (2016) recommends that the chosen number of simulation replications ("runs") should be explained. Hoad et al. (2007), who studied literature on the subject, found that there are three principal ways of determining the number of replications: (1) the rule of thumb; (2) the graphical method, and (3) the confidence interval method, all three offering different levels of tradeoffs between simplicity and accuracy. We started with the first method and ran one thousand replications for each of the twelve scenarios, since Excel can handle the task easily and fast. The standard deviations for the samples obtained from this operation were further used to establish through the confidence interval method that the number of replications used was abundantly sufficient in all cases for the targeted 99 percent confidence interval of one second from the sample means.

It is a methodological expectation for studies using a computer simulation to provide verification and validation of the simulation model (Kelton, 2016). The verification was performed using the suggestions in the work of Kelton (2016): “stressing the code” by making it deal with extreme values of parameters and replacing randomly generated values with those that enable a good prediction of results without using the model. Several iterations were necessary to remove bugs until the model performed perfectly under the verification conditions described above. The simulation model was validated through comparison to prior lift truck research, actual warehouse observations and an interview with a lift truck expert as described in the “Simulation Parameters” section.

FINDINGS

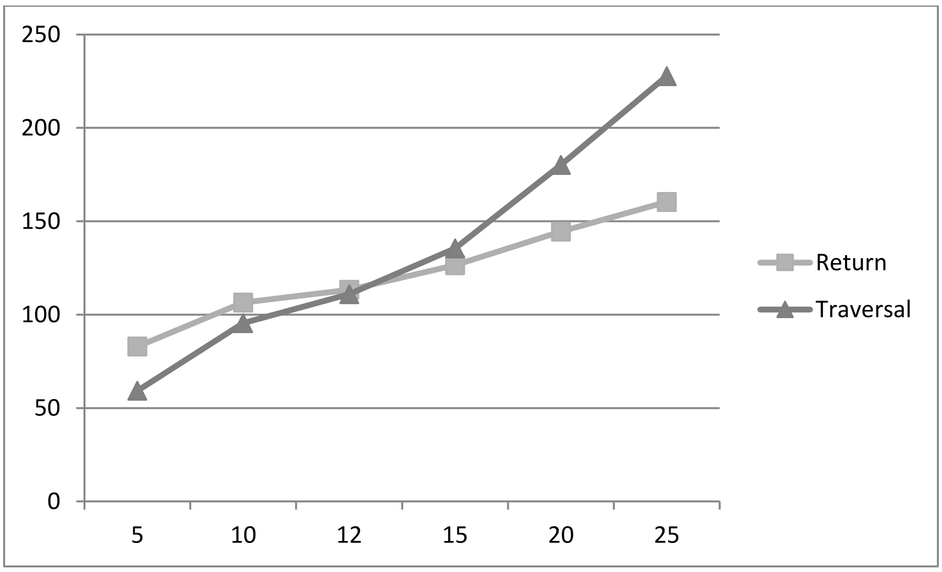
Simulation Results

The simulation results are presented in Table 2 and are described further.

Table 2. Mean travel times for the return and traversal routing policies (in seconds)

POLICY/NUMBER OF PICKS	5	10	12	15	20	25
Return	83	107	113	127	145	160
Traversal	59	95	111	136	180	228

Figure 4. Mean travel times for the return and traversal routing policies (vertical axis, in seconds) for different pick densities (horizontal axis)



The simulation results show that the traversal policy is more efficient (lower travel time) for lower pick densities and the return policy is better for higher pick densities. The point of indifference between the two policies is approximately at the density of 12 picks per 100 locations, or 12 percent. This is graphically represented in Figure 4.

The results provide evidence to support our conjecture that the time penalty for crossing the aisle multiple times will eventually outweigh the inherent advantage of the traversal policy of passing through the aisle only once and become a disadvantage. The simulation made it possible to identify such point.

Comparison to Prior Research

Our findings are dramatically different from the findings for the completely manual picking. For a warehouse with comparable relevant parameters, Goetschalckx and Ratliff (1988b) found that the traversal policy is superior to the return policy for pick densities of up to approximately 65 percent. This effectively means that the traversal policy is better than the return policy for all practical scenarios (Goetschalckx and Ratliff, 1988b), since entirely different solutions are used for high pick densities.

The choice of a routing policy in advance has two implications for the warehouse layout design. Goetschalckx and Ratliff (1988b) pointed out that the traversal policy permits one-way traffic organization, while the return policy requires two-way traffic in the aisle. The one-way traffic organization is also conducive to narrow aisles.

The second implication, which appears to have been completely overlooked by academic research, is the numbering of locations within the aisle. To generate pick lists, most warehouse management systems simply sort the addresses of the picking locations to visit in the ascending order. Locations are traditionally numbered in such a way that odd-numbered locations are on one side of the aisle and even-numbered locations are on the other side of the aisle, as shown in Figure 3. This works perfectly for the traversal policy, which Goetschalckx and Ratliff (1988b) showed to be superior for manual picking. However, in light of our study for lift truck picking, the return policy makes better sense starting from pick densities of 12 percent, but it will not work with numbering locations in this way. Rather, their numbers should increase to the end of the aisle on one side than continue from that point back to the beginning of the aisle on the other side, as shown in Figure 2. This means that sections of a warehouse designated for products with the moving velocity that will likely use the return routing policy should be numbered accordingly.

CONCLUSION

This study has compared the return and traversal routing policies for picking performed entirely with a lift truck in a wide aisle. The study was conducted in a way that allowed a direct comparison to prior research by Goetschalckx and Ratliff (1988b), who considered the same problem for picking performed on foot. In contrast to their study, our results show that the traversal policy is not superior for all practical scenarios but rather only for very low pick densities. Understanding of this point should allow warehouse industrial engineers and managers to select the most beneficial routing policy for a specific scenario and match warehouse layout plans to the picking policies used.

Our study had a number of limitations. While the warehouse scenario we considered was realistic, it was still only one possible scenario. Future research may test the relationship of these and other routing policies in warehouses with different aisle and lift truck parameters as well as consider three-dimensional picking. However, this study was the first to model realistic lift truck movement for routing policies comparison and paved the way for future research in this direction of picking optimization, an increasingly important topic in warehouse operations management.

APPENDIX. TYPES OF MOVEMENT SEGMENTS THAT WERE ESTIMATED

1. Forward linear movement: From stop accelerate to v_{max} or below as distance permits, decelerate to stop. Example: from pos. 1 to pos. 5 (Fig. 3).

2. Forward linear movement: From stop accelerate to v_{max}, maintain v_{max}, decelerate to stop. Example: from pos. 4 to pos. 49 (Fig. 2).

3. Aisle crossing: From stop accelerate to v_1, maintain v_1, decelerate to stop. Example: from pos. 5 to pos. 10 (Fig. 3).

4. Backward linear movement: From stop accelerate to v_1 or below as distance permits, decelerate to stop. Example: from pos. 10 to pos. 8, if location 8 were on the pick list, (Fig. 3).

5. Backward linear movement: From stop accelerate to v_1, maintain v_1, decelerate to stop. Example: from pos. 10 to pos. 6 (Fig. 3).

6. Aisle crossing and continued forward linear movement: From stop accelerate to v_1, maintain v_1, decelerate to stop beyond aisle crossing. Examples: from pos. 6 to pos. 11 or from pos. 11 to pos. 18 (Fig. 3).

7. Aisle crossing and continued forward linear movement: From stop accelerate to v_1, maintain v_1 to end of aisle crossing, accelerate from v_1 to v_{max} or below as distance permits, decelerate to stop. Example: from pos. 6 to pos. 17, if location 17 were on the pick list, (Fig. 3).

8. U-turn outside of the aisle from stopped position: fixed time penalty (Table 1).

Notes: v_{max} – maximum allowed speed (speed limit), v_1 – reduced speed for maneuvering. Items 1 – 7 are applicable to the traversal routing policy; items 1, 2 and 8 only – to the return policy.

REFERENCES

Bowersox, D., Closs, D., Cooper, M., & Bowersox, J. (2013). *Supply Chain Logistics Management*, New York, NY: McGraw-Hill.

Brazhkin V. (2018a). An analysis of key warehouse resources. *Journal of Contemporary Business Issues, 23*(1), 1-19.

Brazhkin, V. (2018b). Outside the box warehousing: When thinking of inputs as outputs makes sense. *Transportation Journal, 57*(2), 219-232.

Burinskiene, A. (2011). The travelling of forklifts in warehouses. *International Journal of Simulation Modelling, 10*(4), 204-212.

Burinskiene, A. (2015). Optimizing forklift activities in wide-aisle reference warehouse. *International Journal of Simulation Modelling, 14*(4), 621-632.

Çelik, M., & Süral, H. (2016). Order picking in a parallel-aisle warehouse with turn penalties. *International Journal of Production Research, 54*(14), 4340-4355.

Dadzie, K. Q., & Johnston, W. J. (1991). Innovative automation technology in corporate warehousing logistics. *Journal of Business Logistics, 12*(1), 63-82.

De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research 182*(2), 481-501.

Faber, N., de Koster, R. (M.) B. M., & van de Velde, S. L. (2002). Linking warehouse complexity to warehouse planning and control structure: An exploratory study of the use of warehouse management information systems. *International Journal of Physical Distribution and Logistics Management, 32*(5), 381-395.

Frazelle, E. H. (2002). *World-class warehousing and material handling*. New York, NY: McGraw-Hill.

Goetschalckx, M., & Ratliff, H. D. (1988a). An efficient algorithm to cluster order picking items in a wide aisle. *Engineering Costs and Production Economics, 13*(4), 263-271.

Goetschalckx, M., & Ratliff, H. D. (1988b). Order picking in an aisle. *IIE Transactions 20*(1), March, 53-62.

Gray, C.F. (1992). An integrated methodology for dynamic labor productivity standards, performance control, and system audit in warehouse operations. *Production and Inventory Management Journal 33*(3), 63-66.

Grosse, E. H., Glock, C. H., Jaber, M. Y., & Neumann, W. P. (2015). Incorporating human factors in order picking planning models: Framework and research opportunities. *International Journal of Production Research, 55*(5), 1260-1276.

Guia, J. T., & Tan Jr., J. B. (2019). Comprehensive analysis of research for warehouse pick path problem. *International Journal of Simulation – Systems, Science & Technology, 20*, July, 20-21.

Hoad, K., Robinson, S., & Davies, R. (2007). Automating DES output analysis: How many replications to run. In *Proceedings of the 2007 Winter Simulation Conference, S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds,* 505-512.

Iyengar, D., Rao, S., & Goldsby, T. (2012). The power and centrality of the transportation and warehousing sector within the US economy: A longitudinal exploration using social network analysis. *Transportation Journal, 51*(4), 373-398.

Kelton, W. D. (2016). Methodological expectations for studies using computer simulation. *Journal of Business Logistics, 37*(2), 82-86.

Kelton, W. D., Sadowski, R. P., & Zupick, N. B. (2015). *Simulation with Arena*. New York: McGraw-Hill Education.

Kunder, R., & Gudehus, T. (1975). Mittlere Wegzeiten beim eindimensionalen Kommissionieren. *Zeitschrift für Operations Research, 19,* B53-B72.

Lee, H. L., Chew, E. P., Frazier, P. I., Jia, Q.-S., & Chen, C.-H. (2013). Advances in simulation optimization and its applications. *IIE Transactions, 45*(7), 683-684.

Masae, M., Glock, C. H., & Grosse, E. H. (2020). Order picker routing in warehouses: A systematic literature review. *International Journal of Production Economics 224*, June, 107564.

McGinnis, M. A., & Kohn, J. W. (1988). Warehousing, competitive advantage, and competitive strategy. *Journal of Business Logistics, 9*(2), 32-54.

McGinnis, M. A., Carlson, R. W., Forry, L., & Fong, L. (1987). Competitive pressures and emerging strategies in public warehousing. *Transportation Journal, 26*(4), 43-53.

Merkuryev, Y., & Merkuryeva, G. (2009). Warehouse order picking process. In *Simulation-based case studies in Logistics.* London: Springer, 147-165.

Murphy, P. R., & Poist, R. F. (1992). Managing the human side of public warehousing: An overview of modern practices. *Transportation Journal, 31,* April, 54-62.

Newsome, K., Thompson, P., & Commander, J. (2013). 'You monitor performance at every hour': Labour and the management of performance in the supermarket supply chain. *New Technology, Work and Employment 28*(1), 1-15.

Peschke, R. E. (2001). Resource capacity and the theory of constraints. In *2001 APICS Constraints Management Technical Conference Proceedings, March 19-20, 2001, San Antonio, TX*, 49-51.

Porter, M. E. (1985). *Competitive Advantage*. New York, NY: The Free Press.

Raney, M. A., & Walter.,C. K. (1992). Electronic data interchange: The warehouse and supplier interface. *International Journal of Physical Distribution and Logistics Management, 22*(8), 21-26.

Richards, G. (2017). *Warehouse management: A complete guide to improving efficiency and minimizing costs in the modern warehouse*. Kogan Page Publishers.

Rouwenhorst, B., Reuter, B., Stockrahm, V., van Houtum, G. J., Mantel, R. J. & Zijm, W. H. M. (2000). Warehouse design and control: Framework and literature review. *European Journal of Operational Research, 122*(3), 515-533.

Stank, T. P., Rogers, D.S., & Daugherty, P. J. (1994). Benchmarking: Applications by third party warehousing firms, *Logistics and Transportation Review, 30*(1), 55-72.

Tompkins, J. A., White, J. A., Bozer, Y. A., Frazelle, E. H., & Tanchoco, J. M. A. (2003). *Facilities planning.* New Jersey: John Wiley & Sons.

DECISION SCIENCES INSTITUTE

Comparative Litigation Rates And Hofstede’s Cultural Dimensions

Nick Heisserer
Central Lakes College
Email: nick.heisserer@clcmn.edu

ABSTRACT

The purpose of this research is to determine what relationship exists between a country's culture and rates of litigation. Understanding this relationship can allow any business entering a new country to better assess its legal risks. A quantitative analysis was conducted between a country’s per capita litigation rates and its cultural dimensions. The results indicated that countries with high power distance dimensions are more likely to have higher litigation rates. More research is recommended in this area to determine if countries not included in this study would reproduce the same relationships.

KEYWORDS: Hofstede, Cultural Dimensions, Litigation Rates, Comparative Analysis

INTRODUCTION

The purpose of this research is to determine what relationship exists between a country's culture and rates of litigation. In this research, litigation rates are defined as civil disputes that include commercial and civil lawsuits but do not include family litigation such as divorce proceedings. Through this research, we hope to be able to better understand what the relationship is between a society’s culture and its litigiousness, and what can be concluded about this relationship if a relationship does indeed exist. As a result of this information, a business that wishes to expand into a new market will be better prepared and informed to conduct business and follow the law.

Litigation is a risk not often considered when a company expands into a new market (Regnier & Tovey, 2007). Causes of litigation can include improper execution of contracts, customer damages, violations of obscure regulations, or investor losses (Barba, 2012). Frivolous lawsuits are a significant risk to the success of an organization because they can create a negative reputation for the company and be expensive to defend in court (Stagnaro, 1997). Studies have demonstrated that advertising campaigns can cause increases in frivolous lawsuits (Gabuthy & Lambert, 2018). The younger the organization, the more significant risk of lawsuits to its growth (Liu et al., 2020). While legal liability can be an area of concern to any organization, it is especially difficult for an organization with intentions to expand into new countries and markets because each country has different legal systems and cultural norms which create variabilities and unknowns. While it can be difficult to understand all the regulations which apply in a new country, it can also be difficult to understand the application of the law. This subjectivity of enforcement is due to cultural differences. Thus, to understand the law, it is important to understand the cultural dimensions of a country.

This research will explore whether Hofstede’s Cultural Dimensions Theory can be utilized as a tool to understand some of this subjectivity. The Theory (Hofstede, 1984) was a seminal work that provided a framework for understanding some of the differences between countries and peoples. Hofstede created this theory as a result of surveys he conducted with IBM employees and their values between 1967 and 1973. Subsequent studies have confirmed his analyses and expanded the number of cultural dimensions from the original four to six. The six cultural dimensions include Power Distance, Individualism, Uncertainty Avoidance, Masculine/Feminine, Time Orientation, and Indulgence. For a detailed description of each dimension, please see Table 1.

Table 1: Description of Hofstede's Cultural Dimensions

Cultural Dimension	**Definition**
Power Distance	Power distance can be defined as the extent to which individuals in a society accept unequal distributions of power. A high-power distance society would expect hierarchy to exist and accept the authority of superiors. A low power distance society would challenge and question those in authority and work to distribute power more evenly.
Individualism	Individualism is the extent to which individuals in society prefer to be integrated into groups. Societies with high individualism scores have weaker ties to large groups in society and are predominately concerned with their immediate families and themselves. Societies with low individualism scores are more collectivist and view themselves as part of a larger social unit in which the needs of the group are more important than the needs of the individual.
Uncertainty Avoidance	Uncertainty Avoidance is the degree to which individuals are comfortable with ambiguity. Societies that score high in uncertainty avoidance create many procedures, rules, or laws to prevent uncertainty with harsh penalties for failure to follow them. A low uncertainty avoidance score indicates more acceptance of the unknown, new, and different. Societies with low uncertainty avoidance impose fewer regulations and are more accepting of disruptions to plan.
Masculine/Feminine	While an outdated use of terms in the 21st century, Hofstede's original meaning of these terms was indicative of the era in which this study was published. Masculine or feminine traits could be applied to both men and women and were more a measure of levels of competition or cooperation. Societies with higher masculine scores place more importance on achievement, assertiveness, and success. Societies with lower masculine scores are considered more feminine which values cooperation and concern for others.
Time Orientation	Time orientation indicates the extent to which society places importance on planning or reacting. A high time orientation score suggests a long-term orientation mindset and a focus on the future. A low-time orientation score prioritizes the importance of the past and the status quo.
Indulgence	Indulgence indicates the amount of freedom a society encourages individuals to use to focus on their own needs and desires. Societies with high indulgence tendencies do not restrict the desires of individuals to do what they please. Societies with low indulgence scores tend to restrain self-gratification.

Research literature has explored a variety of legal applications to Hofstede's Cultural Dimensions Theory (Hofstede, 1984). One study explored the relationship between the transparency of business regulations and the cultural dimensions of individualism, uncertainty avoidance, and time orientation (Gallego- Álvarez et al., 2020). The study concluded that countries with low individualism, uncertainty avoidance, and time-orientation dimensions are less transparent about public laws and regulations. This presents a challenge for a business expanding operations into a new country because it may be unable to understand all applicable regulations and ensure compliance. The decision of what markets to expand into and the method of expansion also appear to be related to cultural dimensions. Businesses prefer to expand into markets that have similar cultural dimensions to their own (Beugelsdijk et al., 2017). However, as markets become saturated, it is not always possible to expand areas with similar cultural dimensions. Greenfield ventures are the preferred expansion method when moving into a country with different cultural dimensions (Beugelsdijk et al., 2017). However, Greenfield ventures are usually more expensive, complex, and subject to legal challenges when the company is not familiar with local regulations (Hand, 2002). In addition to the potential liabilities of any business expansion, COVID-19 has created additional risks. Studies in research literature have suggested that the number of lawsuits may increase after COVID-19 due to employee health damages and the inability to honor contract commitments (Gannon, 2020). These studies suggest that it is important for a business with expansion plans to determine its risks for litigation to create a higher probability of success.

While extant research has documented many independent variables that can affect litigation rates, the root cause of these differences is due to cultural differences. For example, there is a significant difference between the U.S. and Japan regarding the scope of lending services offered by pawnshops. Likely, the country that has more services (such as pawnshop lending) in a sector of the economy will also receive more litigation within this sector. A 1988 study of pawnshops between the U.S. and Japan indicated that in the U.S. 6,853 pawnshops provided 35 million loans per year compared to only 700 pawnshops in Japan that did not provide loans. Consumer preferences were the reason that Japan had fewer pawnshops that did not provide unsecured loans to consumers (Caskey, 1991). Cultural differences can explain these differences in consumer preferences (Mooij, 2004). Another example is divorce litigation. How divorces are administered is another function of culture (Shim, et al., 2013). To again compare the U.S. and Japan, U.S. law requires that couples who wish to divorce must file a suit in court. To divorce in Japan requires the couple to enter a divorce in the family registry at city hall (Ramseyer & Rasmusen, 2010). As a result of this difference, legal proceedings for divorces will be recorded far less in Japan than in the U.S. Finally, per capita car ownership can also be explained by cultural differences. (Sheller, 2004). Countries that have higher per capita rates of car ownership are likely to also record higher rates of car accidents and as a result, higher rates of car accident litigation. (Yannis & Cohen, 2016). These examples demonstrate that culture can be understood to be a root cause for the variation of per capita litigation rates.

Problem Statement:

When deciding to expand internationally, the likelihood of litigation can be an unknown risk and expense to any firm. While much research has been conducted to explore litigation rates and analyze how the cultural dimensions of a country influence other aspects of business, no comprehensive study between countries has been conducted to explore the relationship between litigation rates and culture. Evaluating if a relationship exists and information about this relationship would be beneficial for any organization with interests to expand globally.

Research Question:

What relationship exists between a country's culture and rates of litigation?

LITERATURE REVIEW

A variety of research has been conducted about the causes of litigation rates. Ramseyer and Rasmusen (2010) evaluated litigation rates between the developed countries of Australia, Canada, France, Japan, the UK, and the United States. The authors chose these countries due to their similar amounts of economic wealth and democratic political system to reduce the number of differences between other independent variables that may affect litigation rates of the countries. Litigation rates were measured per 100,000 individuals. Dependent variables included Suits filed, Number of Judges, Lawyers, Motor insurance, and Cost of contract action. Independent variables were the countries measured. Research indicated that the United States legal system had the highest number of lawsuits filed per capita. The authors attributed the high number of lawsuits per capita in the U.S. to frivolous lawsuits.

The article *Judicial Performance and its Determinants: A Cross-Country Perspective* (Palumba et al., 2013) studied legal systems from a variety of countries in the Americas, Europe, and Asia. The article also provided information about lawsuits per capita as well as differences between countries regarding the average length of a trial, computerization within the legal system, and the usage of commercial courts. Data collected about litigation rates was provided as a ratio to the total population of the country. Data were obtained from a questionnaire that was submitted returned from law professionals in the countries included in the study. The study indicated that the diversity of laws and legal regulations were the causes of legal differences between countries.

Cultural Dimensions

Much research has been conducted to evaluate the relationship between Hofstede’s Cultural Dimensions Theory (Hofstede, 1984) and a variety of factors. One study attempted to determine if there was a relationship between Hofstede’s cultural dimensions and obesity (Tekes et al., 2019). The study determined that uncertainty avoidance and individualism were positively associated with obesity and long-term orientation was negatively associated with obesity. The recommendations of the study were to consider Hofstede’s cultural dimensions when developing strategies and campaigns to decrease obesity.

Researchers have also compared Hofstede’s cultural dimensions with innovation. One study evaluated the relationships between Hofstede’s cultural dimensions and corporate innovation policies (Gallego-Álvarez, & Pucheta-Martínez, 2020). The study determined that the cultural dimensions of power distance, masculinity, uncertainty avoidance, and long-term orientation are positively associated with innovation, while the dimension of individualism had a negative effect.

The impact of Hofstede’s cultural dimensions on tourism was also studied. When selecting which informational materials to examine about travel options, the cultural dimensions of power distance, uncertainty avoidance, and individualism within the tourists were positively associated with the selection of specific travel informational resources (Yacout & Hefny, 2014). For example, individuals with high power distance cultures utilized opinions of people with power (such as travel agencies) as important to making their decision. Tourists from countries with high uncertainty avoidance scores were less likely to use friends as a source of information. Tourists from highly collectivistic countries were more likely to use magazines to obtain vacation information.

A study was conducted to evaluate the relationship between cultural differences and whistleblowing. Managers from the U.S. and Croatia were provided with scenarios involving misconduct and interviewed to determine whether they would report the corruption. This data was reviewed in conjunction with the cultural dimension score of each manager (Hofstede, 1984). The study concluded that U.S. managers scored higher on the individualistic cultural dimension than Croatian managers and were more likely to report corruption. Managers who scored higher in masculinity and power distance were also less likely to worry about reprisals (Tavakoli et al., 2003). Croatian managers reported higher cultural dimensions of collectivism than the U.S managers and were more likely to blow the whistle when the needs of the group were placed at harm.

Each previously mentioned study utilized cultural dimension data from specific countries that have been recorded and collected by the Hofstede Institute. The Institute hosts an online data tool at https://www.hofstede-insights.com/product/compare-countries/ that allows public access to the data via the web ("Compare countries," 2017). Data presented by this tool includes all data collected by the institute beginning in 1973 and is updated when new information becomes available. Some information in the tool has been collected as recently as 2020. Users can search for the dominant Power Distance, Individualism, Masculinity, Uncertainty Avoidance, Long-Term Orientation, and Indulgence dimensions of each country. Data for each country are presented in a graph format on a scale of 1-100 with 50 as the median.

Summary of Literature

A review of extant literature about litigation rates and cultural dimensions provides additional context and information about the research question of this paper. Research about litigation rates suggests that the design and controls of the legal system are a strong predictor of the number of litigation rates reported in a particular country. As previously discussed, while several independent variables such as the design of a legal system or its controls can explain the variation of per capita litigation rates, cultural differences are understood to be the root cause for those differences. Because human choices are influenced by their culture and environment (Brislin, 2000), it is hypothesized that the cultural dimensions of a country can be understood to be a primary influencer for the differences in litigation rates among countries. While extant literature reviewed in this paper has also documented a variety of constructs influenced by cultural dimensions, no studies were located to confirm if a relationship exists between per capita litigation rates and cultural dimensions. As a result, research into this area appears to be an original and relevant topic to study.

HYPOTHESIS

Utilizing Hofstede's Cultural Dimensions Theory (Hofstede, 1984), and a review of extant literature, it is hypothesized that cultural dimensions can influence per capita litigation rates. In addition to this primary hypothesis, two secondary hypotheses are proposed:

Secondary Hypothesis 1:

Countries that possess high individualism scores utilizing Hofstede's Cultural Dimensions data will have higher litigation rates compared with countries with high collectivist scores. Societies with elevated levels of individualism are more likely to promote conflict and opportunism (Goncalo & Staw, 2006), so it is likely that more frivolous lawsuits will be brought to court.

Secondary Hypothesis 2:

Countries that possess low uncertainty avoidance scores utilizing Hofstede’s Cultural Dimensions data will have higher litigation rates compared with countries with high uncertainty avoidance scores. Countries that have low levels of uncertainty avoidance are more likely to create fewer laws and procedures to govern behavior. Due to fewer laws, opportunities for litigation will increase because there will be more opportunities for interpretation. According to research, individuals who live in low uncertainty avoidant cultures are more likely to act impulsively (Giebels et al., 2017), which may also increase the possibility of legal reprisals.

METHODOLOGY, POPULATION & SAMPLING:

To evaluate the validity of the primary and secondary hypotheses, a quantitative analysis will be performed that compares litigation rates to cultural dimension scores. The first data set (Data Set 1) will include data from *Comparative Litigation Rates (*Ramseyer and Rasmusen 2010). The report was previously described in the literature review and contains information about per capita litigation rates from a variety of countries. Litigation rates were displayed numerically per 100,000 individuals and will be compared with the cultural dimensions from each respective country. Excluded from litigation rate calculations were criminal prosecutions, and personal proceedings, such as divorce. Countries evaluated were Japan, France, Canada, UK, Australia, and the U.S.

The second data set (Data Set 2) will include data from *Judicial Performance and its Determinants: A Cross-Country Perspective* (Palumba et al., 2013). Per capita lawsuits were calculated numerically as a percentage of the total population and will be compared with the cultural dimensions from each respective country. Excluded from litigation rate calculations were criminal prosecutions, and personal proceedings, such as divorce. Countries evaluated were Finland, Norway, Luxembourg, Sweden, Japan, Israel, Denmark, Austria, Estonia, Poland, Hungary, Switzerland, Slovenia, South Korea, France, Portugal, Germany, Italy, Greece, Spain, Czech Republic, and Russia.

The two data sets are being compared with cultural dimension scores separately because the per capita litigation rates on each are calculated slightly differently. *Comparative Litigation Rates* did not include patent lawsuits in its per capita measures, while *Judicial Performance and its Determinants: A Cross -Country Perspective* did not disclose whether those metrics were included/excluded. Each data set’s per capita litigation rates were compared with the cultural dimensions of each country using the Hofstede Institute’s Country Comparison tool ("Compare countries," 2017). This tool was previously summarized in the literature review. By examining each country’s litigation rates and cultural dimensions together it can be determined whether a relationship exists between the two variables.

Reliability & validity of research variables

The information sets used in this study have been peer-reviewed with sources provided for its findings. Litigation rates from Data Set 1 were obtained from the National Center for State Courts (NCSC). The NSCS receives its data from surveys submitted by legal practitioners and uses statistical analyses to extrapolate per capita litigation by country. Litigation Rates from Data Set 2 were obtained from an OECD questionnaire distributed to Organization for Economic Cooperation and Development (OECD) member and partner countries and compiled with their responses using statistical methods to determine their validity. While no data can be perfectly free from bias, it appears that reasonable precautions have been made to ensure the data presented is accurate and reliable.

RESULTS

Data Set 1: Comparative Litigation Rates (Ramseyer& Rasmusen, 2010) & Cultural Dimension Index ("Compare countries," 2017)

Data for each country was compiled following the parameters of the methodology section. Data were compiled in Table 2. Cell data above the mean was shaded light green, and data that was below the mean was shaded light yellow. Upon review of the data, no relationships were found between per capita lawsuits and cultural dimensions for each country in Data Set 1.

Table 2: Data Set 1 Litigation rates per 100,000 and Hofstede Cultural Dimensions scores

Country	Litigation Rates	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	Long Term Orientation	Indulgence
Japan	1.768%	54	46	95	92	88	42
France	2.416%	68	71	43	86	63	48
Canada	1.450%	39	80	52	48	36	68
U.K.	3.681%	35	89	66	35	51	69
Australia	1.542%	38	90	61	51	21	71
U.S.	5.806%	40	91	62	46	26	68

Data Set 2: Judicial performance and its determinants: a cross country perspective (Palumbo et al., 2013) & Cultural Dimension Index ("Compare countries," 2017)

Data for each country was compiled following the parameters of the methodology section. Data were compiled in Table 3. Cell data above the mean was shaded light green, and data that was below the mean shaded light yellow. Findings suggest that countries with high power distances are positively associated with more lawsuits per capita.

Table 3: Data Set 2 Litigation rates to the total population and cultural dimensions scores

Country	Litigation Rates	Power Distance	Individualism	Masculinity	Uncertainty Avoidance	Long Term Orientation	Indulgence
Finland	0.32%	33	63	26	59	38	57
Norway	0.37%	31	69	8	50	35	55
Luxembourg	0.41%	40	60	50	70	64	56
Sweden	0.65%	31	71	5	29	53	78
Japan	0.72%	54	46	95	92	88	42
Israel	0.89%	13	54	47	81	38	
Denmark	1.32%	18	74	16	23	35	70
Austria	1.34%	11	55	79	70	60	63
Estonia	1.61%	40	60	30	60	82	16
Poland	2.15%	68	60	64	93	38	29
Hungary	2.16%	46	80	88	82	58	31
Switzerland	2.18%	34	68	70	58	74	66
Slovenia	2.24%	71	27	19	88	49	48
South Korea	2.25%	60	18	39	85	100	29
Slovakia	2.37%	100	52	100	51	77	28
France	2.85%	68	71	43	86	63	48
Portugal	2.95%	63	27	31	99	28	33
Germany	3.47%	35	67	66	65	83	40
Italy	3.97%	50	76	70	75	61	30
Greece	4.03%	60	35	57	100	45	50
Spain	4.21%	57	51	42	86	48	44
Czech Republic	4.52%	57	58	57	74	70	29
Russia	9.62%	93	39	36	95	81	20

DISCUSSION

Data Set 2 which utilized the data from *Judicial performance and its determinants: a cross-country perspective* (Palumbo et al., 2013) indicated a possible correlation between countries with a high-power distance and a higher percentage of litigation. According to the data, countries which had per capita lawsuits above the mean were 87.5% were also likely to have a high-power distance score. However, the mean data may have been distorted due to Russia's high litigation rates of more than twice the rate of the next closest country, the Czech Republic. To ensure this significant difference did not distort the results, the median litigation rate was also calculated. The median litigation ratio was Hungary at 2.16%. Utilizing the median, countries at or above the median of per capita lawsuits also had a 77% chance to also have elevated Power Distance scores.

It is not yet discernable whether the primary hypothesis that the cultural dimensions of a country can influence litigation rates. More study is required to confirm or deny this hypothesis. Both secondary hypotheses proposed in this paper were not supported by the research results. However, there was a positive relationship demonstrated between countries with high power distance dimension scores and high per capita litigation rates. Why might societies with high power distance dimensions also report high litigation rates? Initially, it may seem counterintuitive to suggest that societies that accept unequal distributions of power may be inclined to litigate at higher rates. Ramseyer and Rasmusen (2010) had previously theorized that high litigation rates were a symptom of a legal system that created more incentives to bring forward lawsuits than to defend a lawsuit in court. Perhaps countries with high power distance dimensions are more likely to create legal systems that lack the controls to prevent frivolous lawsuits. This may warrant more exploration and further research.

Limitations, and Recommendations

There are several limitations to the scope of these results. The first limitation is the possibility of other independent variables not associated with cultural dimensions which may affect the differences in litigation rates between countries. Specific circumstances and current events, such as a county's political stability, geopolitical conflicts, economic wealth, availability of valuable resources, or historical events may also contribute to litigation rates. However, as previously asserted, extant research has affirmed that culture is a root cause of many types of independent variables (Beraho & Elisu, 2011). As such, this study is still a valid exploration into the relationship between a country's culture and litigation rates. More research into this area is recommended.

Another limitation of this study is that it is difficult to acquire information about per capita lawsuit numbers from every country calculated consistently. This is the reason why two studies were referenced in this research. Not every country maintains this information or has made it publicly available. While the list of countries included in this research was comprehensive, they were not complete. Litigation rates for middle eastern, eastern, and southern hemispheric countries were limited or not able to be retrieved. As a result, the findings of this research cannot be considered a definitive answer to the research question. Even so, these findings do contribute to the field of study related to cultural dimensions and per capita litigation rates. Additional research is recommended to include countries and their litigation rates not included in this study to further explore this relationship.

CONCLUSION

The research question: “*What relationship exists between a country's culture and rates of litigation?” r*equires more data to be definitively answered. However, this study does indicate that a positive relationship exists between the cultural dimension of high-power distance and high per capita litigation rates. More research is recommended to affirm these findings and investigate the causes for this relationship. Correlations do not always affirm causation (Creswell & Creswell, 2018), and it is necessary to explore these issues further so that a business expansion can accurately assess the risks of litigation for any new business venture.

REFERENCES

Barba, R. (2012). Settling Investor Lawsuits Becoming an Expected M&A Cost. The American Banker, 177(F309).

Beugelsdijk, S., Kostova, T., van Essen, M., Kunst, V., & Spadafora, E. (2018). Cultural Distance and Firm Internationalization: A Meta-Analytical Review and Theoretical Implications. Journal of Management, 44(1), 89–130. https://doi.org/10.1177/0149206317729027

Beraho, E., & Elisu, R. (2011). Influence Of Country Culture On Bankruptcy And Insolvency Legal Reform Management. International Journal of Management & Information Systems, 14(2). https://doi.org/10.19030/ijmis.v14i2.829

Brislin, R. (2000). Understanding culture's influence on behavior (2nd ed.). Wadsworth/Thomson Learning.

John P. Caskey, "Pawnbroking in America: The Economics of a Forgotten Credit Market," Journal of Money, Credit, and Banking, 23 (1991) 85-100

Compare countries. (2017, September 4). Hofstede Insights. https://www.hofstede-insights.com/product/compare-countries/

Creswell, J. W., & Creswell, D. J. (2018). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches.* Singapore: Sage Publications, Inc.

Gallego-Álvarez, I., & Pucheta-Martínez, M. (2020). Hofstede's cultural dimensions and R&D intensity as an innovation strategy: a view from different institutional contexts. Eurasian Business Review. https://doi.org/10.1007/s40821-020-00168-4

Gallego-Álvarez, I., & Pucheta-Martínez, M. (2020). How cultural dimensions, legal systems, and industry affect environmental reporting? Empirical evidence from an international perspective. Business Strategy and the Environment, 29(5), 2037–2057. https://doi.org/10.1002/bse.2486

Gannon, John. (2020). COVID Lawsuits. BusinessWest, 36(29), 26–41.

Mark Hand. (2002). The perils of Greenfield Gas development. Public Utilities Fortnightly, 140(2), 28–.

Giebels, E., Oostinga, M., Taylor, P., & Curtis, J. (2017). The Cultural Dimension of Uncertainty Avoidance Impacts Police-Civilian Interaction. Law and Human Behavior, 41(1), 93–102. https://doi.org/10.1037/lhb0000227

Goncalo, J., & Staw, B. (2006). Individualism–collectivism and group creativity. Organizational Behavior and Human Decision Processes, 100(1), 96–109. https://doi.org/10.1016/j.obhdp.2005.11.003

Hofstede, G. H. (1984). *Culture's consequences, international differences in work-related values*. Sage.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations.* Thousand Oaks, CA: Sage Publications.

Liu, C., Cheong, C., & Zurbruegg, R. (2020). Rhetoric, Reality, and Reputation: Do CSR and Political Lobbying Protect Shareholder Wealth against Environmental Lawsuits? Journal of Financial and Quantitative Analysis, 55(2), 679–706. https://doi.org/10.1017/S0022109019000073

Mooij, M. (2004). Consumer behavior and culture: consequences for global marketing and advertising. Sage Publications.

Palumbo, G., Giupponi, G., Nunziata, L., & Mora-Sanguinetti, J. (2013). *Judicial performance and its determinants: a cross-country perspective*. OECD. https://www.oecd.org/economy/growth/FINAL%20Civil%20Justice%20Policy%20Paper.pdf

Ramseyer, M., & Rasmusen, E. (2010). Comparative Litigation Rates. *Harvard Law School*.

Regnier, E., & Tovey, C. (2007). Time horizons of environmental versus non-environmental costs: evidence from US tort lawsuits. Business Strategy and the Environment, 16(4), 249–265. https://doi.org/10.1002/bse.494
Sheller, M. (2004). Automotive Emotions: Feeling the Car. Theory, Culture & Society, 21(4–5), 221–242. https://doi.org/10.1177/0263276404046068
Shim, H., Choi, I., & Ocker, B. (2013). Divorce in South Korea: An Introduction to Demographic Trends, Culture, and Law. Family Court Review, 51(4), 578–590. https://doi.org/10.1111/fcre.12054
Stagnaro, F. (1997). More employee lawsuits raise costs for businesses. San Francisco Business Times, 11(52), 8.a–.
Tavakoli, A.A. Kennan, John P & Crnjak-Karanovic, B. (2003). Culture and Whistleblowing an Empirical Study of Croatian and United States Managers Utilizing Hofstede's Cultural Dimensions. Journal of Business Ethics, 43(1/2), 49–64. https://doi.org/10.1023/A:1022959131133
Tekeş, B., Üzümcüoğlu, Y., Hoe, C., & Özkan, T. (2019). The Relationship Between Hofstede's Cultural Dimensions, Schwartz's Cultural Values, and Obesity. Psychological Reports, 122(3), 968–987. https://doi.org/10.1177/0033294118777965
Yacout, O., & Hefny, L. (2015). Use of Hofstede's cultural dimensions, demographics, and information sources as antecedents to cognitive and affective destination image for Egypt. Journal of Vacation Marketing, 21(1), 37–52. https://doi.org/10.1177/1356766714538444
Yannis, G., & Cohen, S. (2016). Traffic safety. ISTE Ltd.

DECISION SCIENCES INSTITUTE

A design science research approach to continuous process improvement

Benjamin Agyei-Owusu
Kwame Nkrumah University of Science and Technology
Email: bagyei-owusu.ksb@knust.edu.gh

Grace Caroline Fidelia Annan
Kwame Nkrumah University of Science and Technology
Email: fgaanansewaaannan@st.knust.edu.gh

David Asamoah
Kwame Nkrumah University of Science and Technology
Email: dasamoah.ksb@knust.edu.gh

Francis Kofi Andoh-Baidoo
University of Texas Rio Grande Valley
Email: francis.andohbaidoo@utrgv.edu

ABSTRACT

Despite the growth of continuous process improvement as an important quality management and improvement paradigm, publication of process improvement studies in top operations management journals have been very limited. In this study, we propose a design science research approach to guide the conduct, presentation and publication of continuous process improvement projects. A design science-based framework for guiding continuous improvement projects is developed and tested using an on-going Lean Six Sigma quality improvement case study. The developed framework can guide scholars and quality management practitioners to conduct and publish their continuous process improvement projects.

KEYWORDS: Continuous process improvement, Design Science, Frameworks, Lean Six Sigma

INTRODUCTION

Continuous process improvement is a comprehensive and systematic methodology which enables firms to achieve excellence in quality (Sanchez Ruiz et al., 2020). The continuous improvement approach to quality management is directly or indirectly, related to waste mitigation and variability reduction as a way of contributing to enhance the organization's competitiveness (Aqlan and Al-Fandi, 2018). Continuous process improvement techniques such as Six Sigma, lean and kaizen have been adopted extensively to great benefits (Kumar et al., 2018; Ahmed et al., 2019). The objective of continuous improvement studies is usually establishing a culture of sustained improvement (Delgado et al., 2012) and a desire to achieve competitive excellence (Aqlan and Al-Fandi, 2018).

Although continuous process improvement studies and projects has been widely accepted and implemented in practice, very few process improvement-based research studies have been accepted for publication in top operations management journals. Most continuous improvement-based studies have been consigned to practice-based quality journals. The reason may be their lesser focus on theory, a relatively smaller contribution to knowledge offered by continuous improvement-based manuscripts, and a lack of alignment of continuous improvement-based studies to usually accepted presentation formats of most top journals (Peffers et al., 2007). This study aims to address these issues by proposing a design science research-based framework for presenting and publishing continuous process improvement research and projects.

Design science research is an accepted and growing paradigm for conducting research in operations management (Van Aken et al, 2016). Design science focuses jointly on the production of knowledge and actual action on field (Kaipia et al, 2017). We propose that continuous improvement projects can be conducted and presented within the paradigm of design science so that there is a clear contribution to knowledge to add to the practical implementation of continuous improvement project. This would considerably strengthen the continuous improvement-based research projects for acceptance (Van Aken et al, 2016). The study also develops a framework that can guide the conduct, presentation and publishing of continuous improvement-based research and projects within the paradigm of design science research.

The subsequent sections present the background frameworks and theory, after which the proposed framework for continuous process improvement studies is presented and discussed. Next the framework is demonstrated using an on-going case study. The paper then concludes with a conclusion section.

LITERATURE REVIEW

Some key frameworks and philosophies were important in shaping our design science-based framework for continuous improvement studies.

The CIMO-logic framework

Important frameworks like CIMO (Context, Intervention, Mechanism, Outcome) logic have been adopted for conducting design science research in operations management (Groop et al., 2017). There are four stages which are discussed next.

Context describes what the problem entails or the cause of the problem itself, identifying the undesirable effects in the system and understanding the stakeholders. The context also includes challenges and opportunities. *Mechanisms* is the “how” part. This is the evidence of the casualties between interventions and output. understanding the links between the undesirable effect and how it can link to new interventions. *Interventions* is the “what” part. It involves an attempt to solve the problem by resolving conflicts, constructing the evaporative clouding and formulating the design propositions. Finally, *Outcomes* involves anticipating and evaluating intended and unintended intervention outcomes. The outcome stage also involves re-evaluating problem framing in light of unintended consequences (Groop et al., 2017).

Continuous Process Improvement Using Six Sigma

Six Sigma represents the most widely used continuous improvement approach. Six Sigma may be defined as a business strategy used to improve business profitability, to improve the effectiveness and efficiency of all business processes to meet or exceed customer’s needs and expectations (Chakraborty and Leyer, 2013). The Six Sigma methodology, also referred to as the DMAIC framework, represents a well-defined procedure for quality improvement operated on a project basis, that incorporates several techniques and tools, aimed at helping organizations achieve their strategic targets (Hakimi et al, 2018). There are five phases in the Six Sigma methodology.

The Six Sigma process improvement cycle begins with the *Define* phase. Here, a project charter is created which helps to outline details of the project. Additionally, the project team sets out to understand the needs of customers of the process (Antony et al, 2007). In the *Measure* phase, the problem is quantified. Some questions are asked during this stage such as: How does the process currently perform? Or in other words, what is the magnitude of the problem? Measure is critical throughout the life of the project. In the measure phase, the team refines the measurement definitions and determines the current performance or the baseline of the process (Antony et al, 2007). In the *Analyse* stage, the cause of the problem is identified. The cause of the problem and its analysis is often not given enough attention which may result in implementation of solutions which do not address the true root cause of the issues. It is therefore important for process improvement teams to carefully consider potential root causes to problems faced, formulate hypotheses as to why problems are present and then work to either confirm or disprove the hypotheses (Antony et al, 2007). The *Improve* phase is about mitigating the identified root causes of the problem at hand. At this stage, the project team generate improvement solutions and implement them (Antony et al, 2007). Finally, in the *Control* phase, the project team sets out to maintain the gains that have been obtained from the improvement solutions implemented. The project team is also expected to develop a response plan to counter any future drop in process improvement (Antony et al, 2007).

Lean Six Sigma

Lean management aims at reducing the seven types of wastes: Transport, Defects, Over-processing, Over-production, Waiting, Motion and Inventory (Sodhi et al, 2020). Lean and Six Sigma have been described as being related, and yet distinct. Whilst Six Sigma emphasizes more on the identification and elimination of variation through the application of several statistical tools and supporting software, lean management stresses more on reducing or removing of waste through process and value analysis (Sodhi et al, 2020). Given that these two continuous improvement strategies are somewhat closely related, many quality management scholars and practitioners have encouraged that the two strategies be seen as complementary and could be deployed together for potentially greater benefits (Pacheco et al., 2015). Lean Six Sigma is a well-defined procedure for quality improvement that incorporates several techniques and tools, aimed at helping organizations achieve their strategic targets through reducing waste and reducing process variability.

THEORETICAL BACKGROUND

The goal theory is well grounded in the behavioural literature and specifies that difficult yet specific goal leads to better performance (Linderman et al., 2003). According to the goal-setting researchers, “a goal is the object or aim of an action to attain a specific standard of proficiency, usually within a specified time limit” (Locke and Latham, 2002, p. 705). Process improvement projects operate on the basis of setting clear, measurable process improvement goals and following clearly defined techniques and processes to achieve the set goals. The goal theory has been used in prior continuous improvement studies (Linderman et al., 2003). This study contributes to the goal theory by providing a clear framework through which the setting of specific process improvement goals can be achieved.

DESIGN SCIENCE FRAMEWORK TO CONTINUOUS PROCESS IMPROVEMENT

In response to the lack of a clear framework for publishing continuous process improvement projects, a framework to guide the conduct and presentation of continuous process improvement studies is developed and presented. The framework is developed from the integration of the CIMO logic and Six Sigma process improvement methodology (DMAIC).

There are seven stages in the proposed framework, which correspond to the four stages of the CIMO logic and the five stages of the DMAIC methodology. The seven stages of the proposed framework, together with how they link with the components of the DMAIC methodology and the CIMO logic framework are presented in Table 1. The individual stages of the proposed framework are discussed next.

<table>
<tr><td colspan="3">Table 1: Components of the design science framework for continuous process improvement, DMAIC methodology and CIMO logic framework</td></tr>
<tr><th>Design science approach for continuous improvement</th><th>Continuous improvement approach</th><th>CIMO logic framework</th></tr>
<tr><td>Stage 1: Understand the problem and context</td><td>Define phase</td><td>Context</td></tr>
<tr><td>Stage 2: Establish baseline performance</td><td>Measure phase</td><td rowspan="2">Mechanisms</td></tr>
<tr><td>Stage 3: Generate improvement solutions</td><td>Analyze phase</td></tr>
<tr><td>Stage 4: Implement improvement solutions</td><td>Improve phase</td><td>Intervention</td></tr>
<tr><td>Stage 5: Evaluate improvement solutions</td><td rowspan="3">Control phase</td><td rowspan="3">Outcomes</td></tr>
<tr><td>Stage 6: Maintain improved process performance</td></tr>
<tr><td>Stage 7: Document and communicate performance improvement</td></tr>
</table>

Stage 1: Understand the problem and context: This stage of the framework draws on insights from the “Define” phase of Six Sigma methodology and the “Context” stage of the CIMO logic framework. In the first stage of the framework, the problem to be addressed has to be clearly defined and understood in detail. Relevant stakeholders of the problem have to be identified and the researchers or project team must endeavour to understand the problem from their perspective. Relevant continuous process improvement tools and partly from such

as project charter, process maps, SIPOC, and critical to quality can be used to properly understand the problem and context of the study.

Stage 2: Establish baseline performance: This stage draws on insights from the “Measure” phase of the Six Sigma methodology and partly from the “Mechanisms” phase of the CIMO logic framework. In this stage, the researchers or project group members should endeavour to determine the current performance of the process to be improved. It is important to accurately measure current performance of the process as this will serve as baseline to compare the improved performance to. Relevant continuous improvement tools and techniques such as statistical process control, X/Y matrix and various charts can be used to establish the baseline performance.

Stage 3: Generate improvement solutions: This stage draws on insights from the “Analyse” phase of the Six Sigma methodology and partly from the “Mechanisms” phase of the CIMO logic framework. In this stage, the researchers or project group members should aim to uncover the root causes of defects in processes, understand the key process variables that may be linked to defects, and generate solutions that adequately address the identified root causes of poor process performance. Continuous improvement tools and techniques such as root-cause analysis, fish-bone diagrams, 5 whys, solution prioritization matrix, and solution selection matrix can be employed to generate solutions that address the root cause of poor process performance.

Stage 4: Implement improvement solutions: This stage draws on insights from the “Improve” phase of the Six Sigma methodology and the “Intervention” stage of the CIMO logic framework. In this stage the researchers or project group members implement the selected solutions identified in the previous stage. Relevant continuous improvement tools such as quality costing, root-cause analysis can be employed to implement the improved solutions.

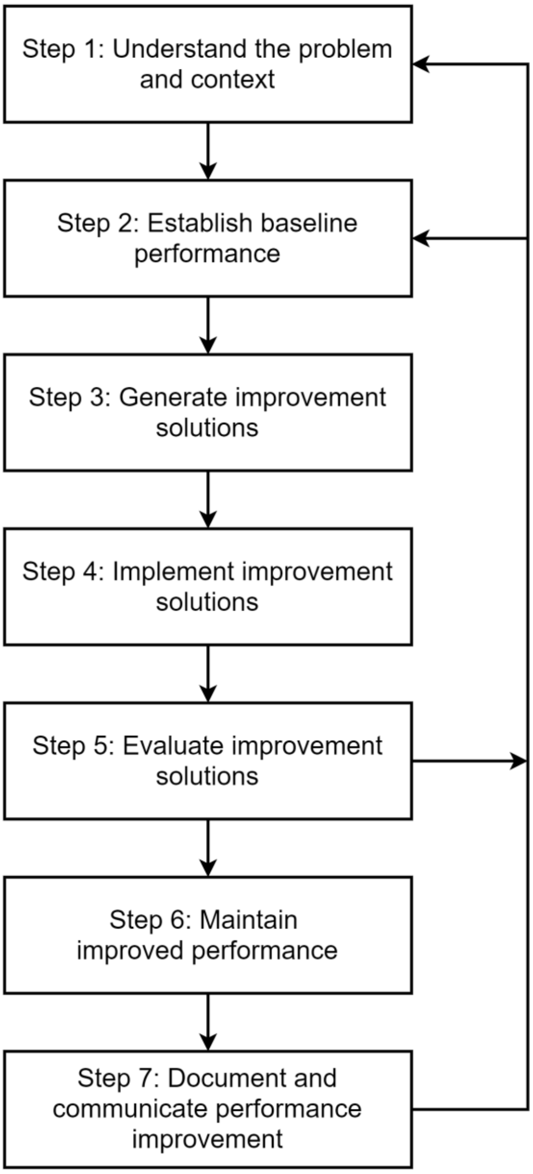
Stage 5: Evaluate improvement solutions: This stage draws on insights from the “Control” phase of the Six Sigma methodology and the “Outcomes” stage of the CIMO logic framework. In this stage, the ability of implemented solutions to meet improvement goals is evaluated in order to validate improvement gains. Continuous process improvement tools such as process capability analysis may be used to evaluate the improvement gains.

Stage 6: Maintain improved process performance: This stage draws on insights from the “Control” phase of the Six Sigma methodology and the “Outcomes” stage of the CIMO logic framework. At this penultimate stage, the researchers or project team members should put in place systems to sustain the improved level of performance. Additionally, control plans should be developed for the new process. Continuous process improvement tools such as failure mode and effects analysis, response planning and control charts can be used to maintain improved process performance.

Stage 7: Document and communicate performance improvement: This stage draws on insights from the “Control” phase of the Six Sigma methodology and the “Outcomes” stage of the CIMO logic framework. At the final stage, the researchers or project group members should seek to document the new methods and processes, identify key lessons from the project, and publish the results internally through company bulletins and externally through publication in conferences and journals.

It should be noted that the seven stages do not necessarily follow a sequential order from stage 1 to stage 7. There may be situations that would require the project group to go back to earlier stages. For instance, if after evaluating the improvement solutions in stage 4, it is observed that process improvement goals were not met, the project group members may have to return to stage 2 to establish current baseline performance and generate new improvement solutions to tackle the root-causes of the quality defect. In some cases, improvement solutions that do not meet improvement goals may be a sign that the project group members did not obtain an adequate understanding of the problem and its context. In such cases, it is necessary for the project group to return to stage 1 to better understand the problem and context so that appropriate solutions can be developed and implemented. It is also important to note that the framework does not necessarily end at stage 7, that is documenting and communicating performance improvement. The proposed framework can be more accurately viewed as a cyclical one. After performance improvements have been obtained, the researchers or project team members can either restart the process to further enhance process performance or target the improvement of another process in the organization. This cyclical approach is captured in both continuous process improvement and the CIMO logic frameworks. The interrelationship between the various stages of the proposed framework is presented in Figure 1 below.

Figure 1: Interrelationships between the stages of the proposed framework



The proposed framework for continuous improvement is next demonstrated with a case exemplar. The case is an on-going study to reduce waiting time at a hospital in Ghana.

CASE EXEMPLAR: REDUCING WAITING TIME IN A REGIONAL HOSPITAL – INITIAL RESULTS FROM AN ON-GOING STUDY

The use of the design science based continuous improvement framework is demonstrated with an on-going case study in a public general hospital located Ghana. Excessive queuing in hospitals in Ghana has been a major problem which negatively affects health delivery. There are reports of patients spending several hours in queues in hospitals when they go for medical care. This is especially true for public hospitals (Adatara and Amooba, 2020). The study is being conducted to help minimize excessive waiting times in the hospital.

Stage 1: Understanding the problem and context

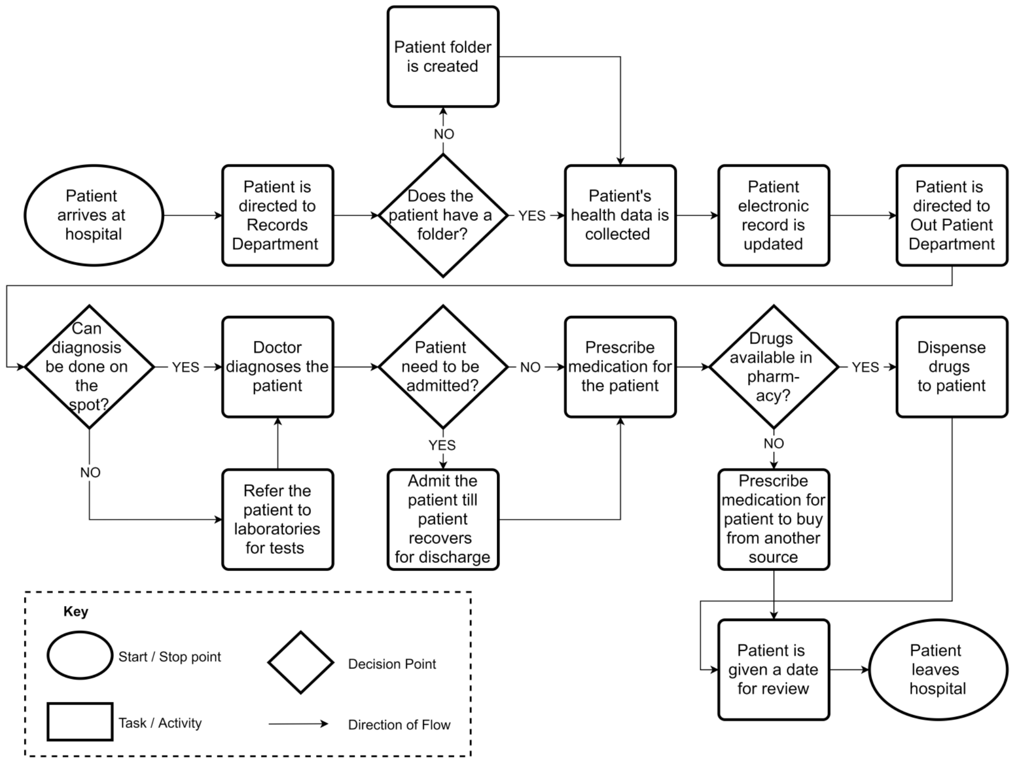
The general problem was that of excessive waiting times for patients who come to seek medical care, which negatively impacts quality of healthcare delivery. Although the hospital has instituted some measures to manage queues such as the institution of “clinic days” when specialized cases are addressed and the adoption of an enterprise resource planning software for digitalization of the hospital processes and administration, there is a still a problem of excessive waiting times. Waiting times are still on the average over 40 minutes. In some instances, patients may spend the whole day at the hospital. Interactions with stakeholders revealed that waiting times of a maximum of 20 minutes was acceptable.

Stage 2: Establishing baseline performance

The researchers interacted with stakeholders in the hospital and developed a process map to help understand the healthcare delivery process from the point patients arrive at the hospital to the point patients leave the hospital. The health delivery process (excluding emergencies) from the point where the customer arrives at the hospital to the point the patient leaves the hospital was mapped and presented in Figure 2 below.

Design science research approach to continuous improvement

Figure 2: As-is Process map



Stage 3: Generating improvement solutions

At this stage, the root causes of the excess delays were investigated and solutions to tackle the root causes were generated. The researchers identified the causes of defects (delays longer than 1 hour) and the waste at all the steps of the healthcare delivery process. The researchers also proposed possible solutions to address the defects. This is summarized in Table 2 below.

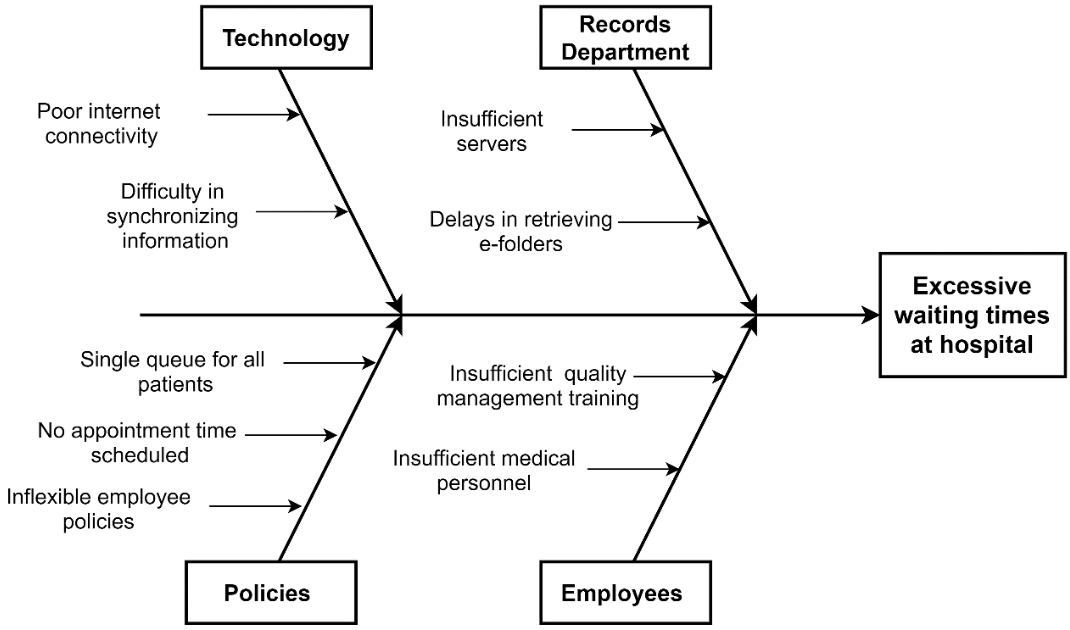
Table 2: Identified defects, wastes and proposed solutions

Activity	**Identified problems**	**Possible solutions**	**Type of waste identified**
Patients arrive at hospital	- Patients with appointments are not given reporting times, only days. - Patient location and traffic situations are not considered. - Patients are served on a first come first serve basis so all patients come early and cause queues.	- Patients must be given specific times to report. - Location and traffic data should be considered in scheduling appointment times.	- Waste of waiting - Waste of motion
Creating and updating patient records	- Misplaced folders. - Network connectivity issues Insufficient servers cause delays	- Patient management software to manage patient records. - Improvement/ upgrade of network connectivity - Get more servers to reduce queuing times	- Waste of motion - Waste of waiting - Waste of overprocessing - Waste of defects (records wrongly updated)
Patient is directed to the OPD	- Patient waiting for long hours to see the doctor.	- More doctors and nurses	- Waste of motion Waste of waiting
Doctor diagnoses the patients	- Insufficient doctors and nurses.	- Flexible doctors and nurses. - Institute overtime	- Waste of waiting
Refer patients to laboratories for tests	- Delays in running the test - Sometimes the tests cannot be conducted by the hospital	- Schedule lab tests. - Ability to link with other health facilities in case lab tests have to be referred	- Waste of motion - Waste of waiting
Prescribe medication for the patients	- Medications are not available	- Real time monitoring of drug availability. - Ability to link with other health facilities in	- Waste of inventory - Waste of waiting - Waste of overprocessing

		case drugs are not available. - Advanced forecasting to keep stock of high demand drugs and less of low demand drugs	
Patient is given a date for review	- Specific times are not given to patients so they end up usually come early to be served first	- Schedule appointments based on time - Schedule appointments based on availability of medical personnel.	- Waste of waiting - Waste of motion

Additionally, the researchers conducted root-cause analysis to better understand what the main causes of the long waiting times in the hospital were. Four broad sources of delays were identified (records department, technology, policies and employees). The root-causes were analysed using a fish-bone diagram presented in Figure 3 below.

Figure 3: Cause and effect diagram



Based on the wastes identified and the root-cause analysis conducted, the following solutions were proposed:

i. Implementing a scheduling management system which incorporates patient location and traffic information
ii. Parallel queues for patients with appointments
iii. Indicating specific times for appointment days
iv. Getting extra servers at the OPD and records section
v. Getting extra medical personnel (doctors and nurses)
vi. Improving the network (internet connectivity) at the hospital

vii. Improving the hospital administrative management system by incorporating location and traffic information

After prioritizing the solutions based on their potential impact on reducing waiting times, implementation difficulty, implementation time, and cost involved in implementation, three solutions were selected:

i. Implementing a scheduling management system which incorporates patient location and traffic information
ii. Parallel queues for patients with appointments and referrals
iii. Getting extra servers at the OPD and records section

Stage 4: Implementing improvement solutions

This is the current stage of the on-going Lean Six Sigma project. The expected outcomes from the implementation of these solutions are:

i. Reduced waiting times to within 20 minutes
ii. Closer match of demand (patients) and supply (doctors and nurses).
iii. Enhanced real time reporting of waiting times to maintain reduction in waiting times.

Stage 5: Evaluate improvement solutions

The performance of the implemented solutions will be assessed to determine if they achieve target performance levels (reducing waiting times to an average of 20 minutes). Depending on the outcome of the evaluation, the project team may either proceed to the next phase or go back to step 1 (understanding the problem and context) or step 2 (establishing baseline performance).

Stage 6: Maintain performance improvement

Response plans will be put in place to ensure that waiting times in the hospital is maintained at 20 minutes on average.

Stage 7: Documenting and communicating performance improvement

Performance improvements (reduction in queue times) will be documented and communicated internally (internal bulletins). Performance improvements will also be will also be published externally (published in journals and conferences).

DISCUSSION AND CONCLUSIONS

This study set out to develop a design science framework to guide the conduct and publication of continuous improvement-based research studies and projects. The proposed framework integrates the Six Sigma DMAIC methodology which is used extensively in continuous improvement studies and projects and the CIMO framework, which is used extensively in design science operations management. The use of the proposed framework was demonstrated by an on-going lean Six Sigma study.

The study has some implications for research and for practice. The components of the proposed framework provide a clear structure that can be adopted to present and publish continuous improvement projects by future researchers. The framework allows future authors to orderly present the activities that occur in continuous improvement projects and encourage authors to make a contribution to knowledge. The proposed framework is applicable for to all continuous improvement studies and future studies can adopt our framework to present their studies in top journals. The study also contributes to the goal theory by providing a clear framework through which the setting of specific process improvement goals can be achieved through continuous improvement projects. Future researchers can adopt the goal theory as a theoretical foundation for their process improvement studies and projects. By way of implications for practice, the study utilized the proposed framework to develop solutions for reducing excessive waiting times in hospitals. These proposed solutions will be implemented to ascertain their efficacy.

Future studies would be directed at completing the implementation of the Lean Six Sigma study presented in the paper so the outcomes can be published. The developed framework needs to be further tested with more case implementations of continuous improvement projects to establish its robustness in presenting continuous improvement project results.

REFERENCES

Adatara, P., & Amooba, P. A. (2020). A Qualitative Study to Explore the Experiences of Older People Utilising Outpatient Healthcare Services from a Teaching Hospital in Ghana. *BioMed Research International*, *2020*.

Ahmed, S., Abd Manaf, N. H., & Islam, R. (2019). Effects of Six Sigma initiatives in Malaysian private hospitals. *International Journal of Lean Six Sigma*.

Antony, J., Downey-Ennis, K., Antony, F., & Seow, C. (2007). Can Six Sigma be the" cure" for our" ailing" NHS?. *Leadership in health services (Bradford, England)*, *20*(4), 242.

Aqlan, F., & Al-Fandi, L. (2018). Prioritizing process improvement initiatives in manufacturing environments. *International Journal of Production Economics*, *196*, 261-268.

Chakraborty, A., & Leyer, M. (2013). Developing a Six Sigma framework: perspectives from financial service companies. *International Journal of Quality & Reliability Management*, 30(3), 256-279

Delgado, A., Weber, B., Ruiz, F., De Guzmán, I. G. R., & Piattini, M. (2014). An integrated approach based on execution measures for the continuous improvement of business processes realized by services. *Information and Software Technology*, *56*(2), 134-162.

Groop, J., Ketokivi, M., Gupta, M., & Holmström, J. (2017). Improving home care: Knowledge creation through engagement and design. *Journal of Operations Management*, *53*, 9-22.

Hakimi, S., Zahraee, S. M., & Rohani, J. M. (2018). Application of Six Sigma DMAIC methodology in plain yogurt production process. *International Journal of Lean Six Sigma*, 9(4), 562-578

Kaipia, R., Holmström, J., Småros, J., & Rajala, R. (2017). Information sharing for sales and operations planning: Contextualized solutions and mechanisms. *Journal of Operations Management*, *52*, 15-29.

Kumar, S., Dhingra, A. K., & Singh, B. (2018). Process improvement through Lean-Kaizen using value stream map: a case study in India. *The International Journal of Advanced Manufacturing Technology*, *96*(5-8), 2687-2698.

Linderman, K., Schroeder, R. G., Zaheer, S., & Choo, A. S. (2003). Six Sigma: a goal-theoretic perspective. *Journal of Operations management*, *21*(2), 193-203.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.

Pacheco, D., Pergher, I., Vaccaro, G. L. R., Jung, C. F., & ten Caten, C. (2015). 18 comparative aspects between Lean and Six Sigma. *International Journal of Lean Six Sigma*.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Sanchez-Ruiz, L., Gomez-Lopez, R., & Blanco, B. (2020). Barriers to effectively implementing continuous improvement in Spanish firms. *Total Quality Management & Business Excellence*, *31*(13-14), 1409-1426.

Sodhi, H. S., Singh, D., & Singh, B. J. (2020). A conceptual examination of Lean, Six Sigma and Lean Six Sigma models for managing waste in manufacturing SMEs. *World Journal of Science, Technology and Sustainable Development*.

Van Aken, J., Chandrasekaran, A., & Halman, J. (2016). Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management. *Journal of Operations Management*, *47*, 1-8.

DECISION SCIENCES INSTITUTE

A Work/Life Teeter Totter: Using Demand Based Principles for Spiritual Operations Management

Dan Bumblauskas, Ph.D.
University of Northern Iowa
Email: daniel.bumblauskas@uni.edu

Paul Bumblauskas
PFC Services, Inc.
Email: paul@pfcservicesinc.com

ABSTRACT

The purpose of this article is to detail how spirituality is a critical component to work/life balance and how individual and organizational spirituality impact operations management. Balancing key areas of work and professional life become an important part of creating happy and mutually satisfying relationships. We have identified six critical elements of life balance that lead towards optimized productivity (in no order of significance): Career, Family, Community, Spiritual, Health (includes sleep), Pleasure. Spirituality is one of the six components that must be defined, with a corresponding work and personal development plan, at an individual and organizational and operations management level.

KEYWORDS: Work-life balance, workplace spirituality, religion / faith, competitive advantage, demand-based approach, family / private businesses

INTRODUCTION

The purpose of this article is to detail how spirituality is a critical component to work/life balance and how individual and organizational spirituality impact workplace culture and influences the consumer marketplace. Previous research has shown that relationships are paramount to happiness. Balancing key areas of work and professional life become an important part of creating happy and mutually satisfying relationships. We have identified six critical elements of life balance that lead towards optimized productivity (in no order of significance): Career, Family, Community, Spiritual, Health (includes sleep), Pleasure. We suggest this is both an analytical and subjective process. Since time is finite, it should be balanced based on the relative importance to the individual and their stakeholders.

Spirituality is one of the six components that must be defined, with a corresponding work and personal development plan, at an individual and organizational [family] firm level. Organization that are closely held, whether family businesses or not, tend to have better control over the influence of spirituality of the drivers of the customer experience. This is because public and larger companies are directed by laws and regulations that cause compliance to be a priority over choice in terms of how spirituality is conveyed as part of the presentation by the entity to its customers and others. Using a demand-based or demand driven modeling approach, there can be a better understanding of what level of spiritual engagement the organization should strive to achieve with stakeholders and customers. When there is a match between the individual (the focal point of the demand-centric view) and organization, the synergy can create a company-defined competitive advantage, assuming this is also a match with the customer and end-user desires. Since the customer/end-user is the focal point of the demand-channel for the entity, a match between the stakeholders, employees and the customers/end-users will allow the organization to trend toward optimal results in terms of market penetration and satisfaction.

BACKGROUND

A team of academic researchers submitted a symposium proposal for The Academy of Management (AOM) 2019 Annual Meeting on *the role of religion in the creation of competitive advantage in family firms*. This topic has spawned various theories, hypotheses, research questions, case studies, and a journal special issue in an attempt to further understand the factors influencing the role of religion, which is a subset of the broader umbrella of spirituality, in creating competitive advantage. The difference between spirituality and religion is well documented as detailed further in the literature review below. There are anecdotal case studies of successful organizations that put religious and spiritual beliefs as a high priority and, in doing so, have influenced the competitive landscape of their industries. However, conversely, there are other cases in which firms have pushed religious and spiritual beliefs to the dismay and outrage of some audiences and consumers.

Whether spirituality and religion are catalysts, have any correlation, etc. in the competitive advantage of firms is highly debatable at least partly dependent on consumer demographics (i.e., how many customers exist within the same or similar spiritual belief structure). Spirituality is also just one of many aspects in which organizations can create competitive advantage, some of which have been previously studied, e.g., operations management, quality, sustainability, corporate social responsibility, etc. There is perhaps some alignment between these seemingly disparate areas, what we describe as the spirituality piece of the “teeter-totter” of work/life balance. From a demand-based perspective there must be synergy between the entity and the party they are attempting to influence for this to be a valuable tool. Spirituality, in and of itself, can be a catalyst, neutral or a hinderance. The primary scope here is family-run and privately held businesses that have fewer secular constraints than publicly traded companies do.

RESEARCH MOTIVATION

Over many years consulting in industry, a model was developed that accounts for six areas necessary to properly balance work and life as shown in table 1. One of those six areas is spirituality. Good health creates quality time, to be managed, and quality time is a pillar to successful management of resources and happiness. Longitudinal research has shown that relationships (personal, professional, spiritual, etc.) drive happiness (Mineo, 2017), so it comes to reason that we should allocate time to this area to create, perhaps, competitive advantages (for ourselves, our employees, our investors, our customer/end-users, etc.).

Table 1. Six Key Areas of Balancing Time and Stakeholder Buy-In

Six Key Areas for Work/Life Balance
A. Career B. Family C. Community

D. Spiritual
E. Health (including sleep)
F. Pleasure

Notes:
These six elements are not listed by priority and should be validated with stakeholders

In industry practice, it has been observed that the role of inputs, such as spirituality and/or religious beliefs, are a key area that should be considered when it comes to family businesses and their productive output of products and services. Therefore, our primary focus in this article is on the need to balance the spiritual component of our time, with a secondary focus on the impact of family business competitiveness. The spirituality component of the models and modules which have been developed for and within organizations must be differentiated from religion, Western or Eastern (Fernando and Jackson, 2006; Rastgar et al., 2012) and optimized in order to build successful personal and professional relationships. In doing so, competitive advantages or disadvantages, can be achieved.

Industry case studies have exemplified this and, in some cases, have been very controversial and polarizing, but many organizations and firms often consciously or unconsciously disseminate their spiritual beliefs in their governance structure and along to stakeholders and consumers. We define stakeholder(s) as a person or persons with an interest or concern to the individual balancing their time and efforts. It is important that there be a match between selling organizations, employing organization, etc. on a spiritual level. Once that match is achieved, the sum of the individual contributions towards the organization will create a competitive advantage as the entity defines it. This can also increase employee and end-user satisfaction. This is an analytical process that is based on subjective priorities. A spreadsheet or time management software can assist with this effort. Since we suggest a demand-centric approach, this means the individual is the focal point and they choose their stakeholder(s). This may include the business as a priority but need not. This also need not be religion, but religion is a common spiritual element.

The definitions of spirituality vs. religion are important considerations (Benefiel, 2003a). Spirituality can be defined a number of different ways, but we have focused on some of the often-cited definitions in our literature review below to validate and verify our interpretation. For our purposes, spirituality is a higher level personal and qualitative life purpose (e.g., macro-level or ‘37,000-foot view’).

It is fairly easy for non-public/non-regulated organizations to achieve consistency between ownership, management, other employees and even vendors & other stakeholders with respect to spirituality broadly and religion more specifically. After all, the ownership often chooses those that have similar spiritual views. However, there is at times a disconnect between the organization and its customers as they pressure a certain view with the assumption that there is an adequate market with similar view that will be able to optimize the value of the organization. Ownership imparting their spiritual beliefs upon employees, especially in family businesses is risky for operations. In addition, this is very supply-side focused, i.e., pushed on employees, rather than demand-side focused. E.g., we will start the meeting with a prayer without discussing if everyone finds that to be acceptable. Alternative approaches might ask if the customer would like to start a meeting with a prayer of their choice and/or the owner or senior management team could ask if everyone is comfortable with a prayer to kick-off the meeting.

Demand Based Spiritual Management in OM

Unless the organization adopts a demand-based view, the organization will not be able to anticipate the outcome of the spiritual pressure that is directly or indirectly applied. An example of direct pressure would be an organization, such as the controversial Chick-fil-A, that screens employees and possibly restricts hiring based on certain spiritual beliefs and the employee boldly sharing their beliefs to the public. This legal and constitutional rights argument is one that must be considered. Indirect pressure is a by-product of the uniform spirituality that is conveyed by comments such as “have a blessed day”- in other words it is tolerated by the organization typically because the organization may not have a balance of various types of religions and/or spiritual beliefs. Further there is a very narrow line between religion and certain secular evolutional developments- an example would be same-sex relationships and marriages. While these are gaining support in industry, spiritual and religious organizations sometimes correlate this with some level of consistency with their religious/spiritual views.

Should organizations place their spirituality aside in order to optimize results and the value of the organization or should the organization be a conduit for spiritual/religious change? This question is similar to questions faced by eco-entrepreneurs as these build businesses and organizations focused on sustainable operations. Today it seems that there is a quiet battle to decide what place spirituality has in business. Because family businesses are the ones that are often at the front lines, because they need not comply with all the regulatory pressures, they seem to be waging the battles in the field. Hence, they are the focal point of this research as it relates to competitiveness and spirituality. Another way to view this is that entities that focus on spirituality value spirituality over other objectives. From a demand-based perspective this needs to be validated as the correct approach given the product or service offering being made available. In some cases, the result will be positive, as least from the perspective of those that agree with the approach. However, the result can be neutral or negative- it depends on how the customer/end-user defines their own work life balance.

We suggest a simple approach. First, determine the work/life balance of the owner(s) or key influencers to the closely held or family business. If spirituality is important then step two is to determine if the demand channel has the same or at least a similar interest in spirituality. If they do, then spirituality influenced approaches can have a positive outcome. Third, structure the message and approach based on the customer/end-user view and the influencer’s view. Lastly, monitor results to assure that the results are a match with expectations. We caution that having spirituality high on the owner/influencer priorities in no way guarantees that spirituality initiatives will have a positive outcome. This needs to be validated from a demand-based perspective and preferably before the spirituality-based initiatives are activated.

LITERATURE REVIEW

Wong and Vinsky (2008) document the separation of spirituality and religion in the literature and applying their methods in the field of social work. They rely heavily on the work of Canda and Furman (1999) who define spirituality and religion as very different things. Other contemporary definitions seemingly reinforce the definitions from the Wong and Vinsky (2008) and Canda and Furman (1999) studies (Divine Intelligence Institute, 2017; Sheldrake, 2009). Others, such as Deepak Chopra, have popularized spirituality in the mainstream media and society.

The University of British Columbia (UBC) Centre for the Study of Human Evolution, Cognition and Culture (HECC) Cultural Evolution of Religion Research Consortium (CERC, https://hecc.ubc.ca/), Heinrich et al. (2010), and Iannaccone (1998) have laid the foundation for much of the work in this area, some of the latter more-so in the area of religion and economics.

Nadesan (1999) looked more closely at religion and corporate capitalism while Conlin (1999), Overell (2003), and Porth et al. (1999) are cited for originally connecting religion with organizational performance and competitiveness. Others have expanded upon this foundation (Dyck, 2014; Tracey, 2012) and Tracey et al. (2014) note the lack of research in this area despite religion's prevalence in societies. It is actually amazing that the field of religion and spirituality in organizational performance is so young compared to the long history of organized religion. Epstein (2002) also notes the need for more consideration of religion in management education, others have aligned religion history and business ethics (Bay et al., 2010; Graafland et al., 2007), and Dyck and Wiebe (2012) consider salvation and organizational management theory.

Fernando and Jackson (2006) conducted an interesting study in Sri Lanka looking at the impact of religious connection in management decision making outcomes (i.e., what impact did religious connectivity have) while Rastgar et al. (2012) did a similar study in Iran. Benefiel (2003a, 2003b, 2005) and Benefiel et al. (2014) examine various aspects of spirituality and organizational research and science. Dent et al. (2005) make use of a robust review of the literature to more deeply explore the relationship between spirituality and organizational / workplace leadership and Fry and Matherly (2006) leadership transformation. Balog et al. (2014) look at the links between entrepreneurship, religion, and spirituality. In the end, all of the literature seems to suggest the need for further theory development and empirical research in this area. The concepts and theories in this paper are based on many years of management consulting in industry and academic research work.

METHODS AND RESULTS

The model framework and case assumption we are making is that family firms are not public, not covered by discrimination laws, etc. Individuals and family firms focused on balancing spirituality, therefore, align with the definition and assumptions provided in previous sections. Understanding the demand channel is key to all of this (Bumblauskas, Bumblauskas, Sapkota, 2017). For example, if a firm wished to promote their spiritual or religious beliefs in the marketplace, to stakeholders, etc. demand chain management would focus on consumers requesting or 'demanding' that the firm exude spiritual or religious principles and foundations in their business operations. The question here is whether companies force religion and spiritual aspects into the marketplace or if this is being driven or requested by the consumer.

One research question is whether spirituality and religion could be used to create a competitive advantage or disadvantage – technically, these elements could be perceived by consumers to be neutral, positive or negative. There seems to be the notion that perhaps there is some form of isolation of mutual exclusivity present with spirituality, religion and competition. As some businesses may choose to focus on spiritual or religious beliefs, similarly some organizations may focus on environmental sustainability, which businesses would want to leverage in order to create a competitive advantage by utilizing spiritual themes. These themes will then yield a positive, neutral or negative reaction from the consumer. To further expand upon the example of a restaurant or consumer goods store that promotes spiritual and/or religious sentiments, boycotts of the firm could ensue. This situation would then include those opposed to the boycotts that support the firms' spiritual or religions foundational principles, those 'grey-area' consumers who are indifferent to boycotts or support of the firm, and the detractors of the firm who are adamantly opposed to the spiritual and religious messages.

Our conceptual belief or hypothesis is that management teams and ownership should not focus on using religion, solely, to create a competitive advantage, but rather communicate their spiritual views and positions as they see fit (by first creating a balance around the six key areas, including spirituality, as listed in table 1). This also reigns true with sustainable practices across countries of operations – e.g., being focused entirely or solely on environmental sustainability may or may not align with consumer *demand-based drivers of success* (Bumblauskas, Bumblauskas & Rosol, 2018; Bumblauskas, Bumblauskas & Sapkota, 2017). Demand driver identification can be further accomplished by aligning supply channel management with the consumer expectations, creating positive benefit via demand channel management. In any case, you may have certain spiritual and/or religious denominations or groups present which can be polarizing and distracting to business enterprises focused on creating competitive advantage for their firm.

Revisiting our restaurant or consumer goods retailer case, the ultimate question could be whether their customers like the products/services, spiritual/religious messaging, the scripted customer service operations protocols (e.g., "have a blessed day"), and/or the combination of themes used by the firm. It is also important to consider how this affects hiring and whether certain employees want to, or do not want to, work for an organization based on their spiritual and/or religious beliefs and affiliations. I.e., will top employees work for the firm(s)? We have witnessed some firms so embattled in the United States court system over such issues and positions that it arguably creates a competitive disadvantage (Marshall, 2012; Morrison, 2013). This may be in the form of a societal and legal disadvantage rather than an overall financial disadvantage.

DISCUSSION AND CONCLUSION

Our general belief is that you need to balance across our six key framework areas, which include spirituality and/or secular business practices. We also believe organizations need to avoid unilateral decision making, for example by consulting only management rather than making potential rash and uninformed decision without consulting stakeholders (specifically customers using the demand-based approach). Examples of this could be embarking on hasty military deployments, going to church on specific days of the week 'or else,' etc. without getting buy-in from demand-based constituents. Individuals and businesses should be able to prioritize and evaluate the highest return on investment activities based on their own criteria. This can lead to potential conflicts, such as how do we define what is a competitive advantage and how does it align with spirituality and religion?

One question that remains to be answered is whether religion, solely, creates a competitive advantage in family firms, and we are interested in the further development of this research line to prove or refute this hypothesis. In this case, religion must be a demand driver so the correlation would tend to be coincidence. If the company only monitors positive outcomes, incorrect conclusion can be reached. There are cultural overlaps as well that often align with religion such as treating people fairly. For example, does controlling your medical service offerings based on beliefs really provide a competitive advantage or hindrance to business and management? In one recent case detailed by Newhoff and Jamison (2018), agnostics vs. Catholics might view this very differently in terms of the services being offered and not offered to customers, or in this healthcare case, patients. Can businesses create a competitive advantage by switching their religious positions, beliefs, and affiliations? Religion is one subset of spirituality in our model. Another example would be individual people from different religious organizations coming together as a couple and their need to balance their spiritual, personal and professional activities (e.g., Jewish and Catholic individuals marrying).

Here we have highlighted some examples from models, management consulting, and industry case experience. More research should be conducted to validate and verify the conceptual theories and hypotheses presented herein. This is a challenging and complicated topic which tends to 'strike a chord' emotionally in a positive or negative way, with consumers and readers. At the end of the day, individuals, organizations and firms need to decide how to best balance the spirituality "teeter-totter" of life balance and how it impacts their operations management strategy.

REFERENCES

Balog, A. M., Baker, L. T., & Walker, A. G. (2014). Religiosity and spirituality in entrepreneurship: a review and research agenda. *Journal of Management, Spirituality & Religion*, 11(2), 159-186.

Bay, D., McKeage, K., & McKeage, J. (2010). An historical perspective on the interplay of Christian thought and business ethics. *Business & Society*, 49(4), 652-676.

Benefiel, M. (2003a). Mapping the terrain of spirituality in organizations research. *Journal of organizational change management*, 16(4), 367-377.

Benefiel, M. (2003b). Irreconcilable foes? The discourse of spirituality and the discourse of organizational science. *Organization*, 10(2), 383-391.

Benefiel, M. (2005). The second half of the journey: Spiritual leadership for organizational transformation. *The leadership quarterly*, 16(5), 723-747.

Benefiel, M., Fry, L. W., & Geigle, D. (2014). Spirituality and religion in the workplace: History, theory, and research. *Psychology of Religion and Spirituality*, 6(3), 175.

Bumblauskas, D., Bumblauskas, P., & Sapkota, K. (2017). "Is Demand Chain Management the New Supply Chain Management? Will the Demand Channel Trump the Supply Channel?" *Journal of Strategic Innovation and Sustainability*, 12(1), pp. 30-40. Available: http://pfcservicesinc.com/?page_id=845.

Bumblauskas, D., Rosol, S., & Bumblauskas, P. (2018). "Managing Multiple Projects: Applying a Demand-Based Approach." *International Journal of Management Education*, 16(1), pp. 52-62. DOI: https://doi.org/10.1016/j.ijme.2017.11.004.

Canda, E. R. and Furman, L. D. (1999) Spiritual Diversity in Social Work Practice: The Heart of Helping, New York, The Free Press.

Conlin, M. (1999). Religion in the workplace: The growing presence of spirituality in corporate America. *Business Week*, November 1: 150-58.

Dent, E. B., Higgins, M. E., & Wharff, D. M. (2005). Spirituality and leadership: An empirical review of definitions, distinctions, and embedded assumptions. *The Leadership Quarterly*, 16(5), 625-653.

Divine Intelligence Institute (2017). "The Difference between Spirituality and Religion." *Divine Intelligence Institute*. Available: https://divineintelligenceinstitute.com/2017/11/15/the-difference-between-spirituality-and-religion/

Dyck, B. (2014). God on management: The world's largest religions, the "theological turn," and organization and management theory and practice. *Religion and Organization Theory* (pp. 23-62). Emerald Group Publishing Limited.

Dyck, B., & Wiebe, E. (2012). Salvation, theology and organizational practices across the centuries. *Organization*, 19(3), 299-324.

Epstein, E. M. (2002). Religion and business–the critical role of religious traditions in management education. *Journal of Business Ethics*, 38(1-2), 91-96.Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, H.C., Bolyanatz, A., Camilo Cardenas, J., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F.W., Tracer, D., and Ziker, J. (2010). "Markets, religion, community size and the evolution of fairness and punishment," *Science* 327: 1480-1484.

Fernando, M., & Jackson, B. (2006). The influence of religion-based workplace spirituality on business leaders' decision-making: An inter-faith study. *Journal of Management & Organization*, 12(1), 23-39.

Iannaccone, L. R. (1998). "Introduction to the Economics of Religion." *Journal of economic literature*, 36(3), 1465-1495.

Marshall (2012). "Chick-Fil-A Takes Families Under Its Wing." *Heritage Foundation*. Available: https://www.heritage.org/religious-liberty/commentary/chick-fil-takes-families-under-its-wing

Mineo, L. (2017). "Good genes are nice, but joy is better." *Harvard Gazette*. Available: https://news.harvard.edu/gazette/story/2017/04/over-nearly-80-years-harvard-study-has-been-showing-how-to-live-a-healthy-and-happy-life/.

Morrsion, M. (2013). "Chick-fil-A Quick to Head Off Potential Gay-Marriage Controversy." *AdAge*. Available: https://adage.com/article/news/Chick-fil-A-quick-head-potential-gay-marriage-controversy/242896/.

Nadesan, M. H. (1999). The discourses of corporate spiritualism and evangelical capitalism. *Management Communication Quarterly*, 13(1), 3-42.

Newhoff, N. and Jamison, T. (2018). "Covenant to stop performing vasectomies, tubal ligations." *Waterloo Cedar Falls Courier.* https://wcfcourier.com/news/local/update-covenant-to-stop-performing-vasectomies-tubal-ligations/article_4884db70-4961-5afb-8151-4917511aae4a.html.

Overell, S. (2003). Spirituality in business: A reconnection with core values. *Financial Times*, 11 (20 September 2003).

Porth, S. J., McCall, J., & Bausch, T. A. (1999). Spiritual themes of the "learning organization". *Journal of Organizational Change Management*, 12(3), 211-220.

Rastgar, A. A., Zarei, A., Davoudi, S. M. M., & Fartash, K. (2012). The link between workplace spirituality, organizational citizenship behavior and job performance in Iran. Arth Prabhand: *A Journal of Economics and Management*, 1(6), 51-67.

Sheldrake, P. (2009). A brief history of spirituality. John Wiley & Sons.

Tracey, P. (2012). Religion and organization: A critical review of current trends and future directions. *The Academy of Management Annals*, 6(1), 87-134.

Tracey, P., Phillips, N., & Lounsbury, M. (2014). Taking religion seriously in the study of organizations. *Religion and Organization Theory* (pp. 3-21). Emerald Group Publishing Limited.

Wong, Y. L. R., & Vinsky, J. (2008). Speaking from the margins: A critical reflection on the 'spiritual-but-not-religious' discourse in social work. *British Journal of Social Work*, 39(7), 1343-1359.

DECISION SCIENCES INSTITUTE

Heterogeneous Uncertainty in the Retail Industry During COVID-19: A Panel Data Exploration

Micah J. Marzolf
Michigan State University
Email: wardmica@msu.edu

Simone T. Peinkofer
Michigan State University
Email: simonep@broad.msu.edu

Jason W. Miller
Michigan State University
Email: mill2831@broad.msu.edu

ABSTRACT

The onset of COVID-19 increased uncertainty across supply chains, with a particularly large impact on retailers due to factors including increased unemployment, altered shopping patterns, and supply disruptions. We employ archival data consisting of 8,647 analyst sales forecasts for 76 U.S. public retailers and manually collected data from press reports and other online sources to explore the impact of COVID-19 on the evolution of sales forecast uncertainty and mean forecasts early in the pandemic, focusing on how the evolution differed for essential versus nonessential retailers. We discuss implications for retailers and contributions to retail operations and economic uncertainty literatures.

KEYWORDS: Retail operations, Economics, Archival research, Econometrics, Panel studies

DECISION SCIENCES INSTITUTE

Increasing Enrollment of Ultra-Orthodox Students in STEM Education: A Case Study about Women Colleges of Engineering in Jerusalem

Michael Naor
Hebrew University of Jerusalem
Email: michael.naor@mail.huji.ac.il

Gavriel D. Pinto
Azrieli College of Engineering
Email: gabipi@jce.ac.il

Pini Davidov
Azrieli College of Engineering
Email: pinida@jce.ac.il

Julie Benaroch
Azrieli College of Engineering
Email: teboul.julie@gmail.com

Adit Kivel
Hebrew University of Jerusalem
Email: kivel@g.jct.ac.il

ABSTRACT
This study examines how the establishment of an ultra-orthodox women program tailored for their need fosters female students' enrolment in STEM and decrease dropout from college. The methodology used is a survey accompanied by rigor case study analysis of the Jerusalem College of Engineering. The orthodox community in Israel represents a unique family cell setting in which the men are obligated to devote workday for Bible studies while women are breadwinners. Therefore, the colleges serve a national priority helping ultra-orthodox households in Jerusalem to sustain which ultimately strengthens Israel's economy.

KEYWORDS: Ultra-Orthodox, Higher Education, Women in Science Technology Engineering and Mathematics, Gender Diversity, Jerusalem, Israel

INTRODUCTION

As part of an attempt to increase inclusion of diverse population on campus, religious students have started receiving a long overdue scholarly attention (Bowman and Smedley, 2013; Cole and Ahmadi, 2003; 2010; Hartley, 2004; Mooney, 2010; Pascarella and Terenzini, 2011). The unique challenges facing ultra-orthodox religious students who hail from fundamentalist denominations, minority groups and sectarian religious communities require a rigor research test to comprehend their motivations, concerns, and experiences in pursuing higher education and to best support them in this journey.

Israel's colleges and universities aren't graduating out engineers and software alumni at the pace the high-tech industry needs. A recent forecast by the Israel Innovation Authority estimated the labor shortfall will reach more than 10,000 over the next decade. Ultra-Orthodox women (Haredim in Hebrew) is a population group that government officials are targeting to fill the shortfall in STEM education. Energized by the growing motivation of Haredi young adults to join the Israeli workforce, and propelled by government need, there has been rapid growth in the bundle offering of academic tracks for the Haredi population since the turn of the millennium. In 2012, in addition to affirmative action stipends offered to Haredi students in Israeli universities, fourteen new government funded academic tracks were established for Haredi students (Golan and Fehl, 2020). The ultra-Orthodox community has been playing a minor role in Israel's flourishing innovation ecosystem. A joint survey articulated by the Israel Advanced Technology Industries (2020) with Haredi employment NGO KamaTech, indicated that in 2018, ultra-Orthodox workers account for about 3% percent of Israel's tech workforce. According to the report, ultra-Orthodox Jews employed in the country's tech sector had climbed 52 percent in the years 2014 to 2018, with women leading the way (Of the 9,700 Haredi tech workers that made up the industry, 6,900 of them were women). Initiatives are made to address this troubling statistic. For example, Start-Up Nation Central affiliate, Scale-Up Velocity, created Adva, a pilot program that trains female students enrolled in religious seminaries to become senior software engineers and website developers. Eighty-one women were part of the first group to graduate from the two-year program. By the program's graduation event date in December 2020, 13 had already secured jobs at tech firms such as Apple, Facebook, XM Cyber, and Check Point Software Technologies.

The statistics mentioned above is evident in orthodox communities were family tradition pressures women to get married in young age and assume housekeeping roles. Therefore, this study about increasing the percentage of women and specifically orthodox in academia have potential to create social change in Israel society: by Providing orthodox (Haredim) with the necessary tools and skills to work as equals in Israel's competitive workforce and become contributors to the economy. This outcome may reduce dependence on government and social welfare because salary earnings in the Israeli tech sector today are 2.5 times greater than in the general economy. Interestingly, nearly three-quarters of ultra-Orthodox women work, an identical figure to the general population, but their average monthly salary of 6,170 shekels ($1,736) is below the 9,309 Shekels average of other Jewish women, and many ultra-orthodox women work only part time.

The current study has several agendas. First, it reviews specifically at Israel the statistical composition of orthodox women vs. non-orthodox women studying Sciences, Technology, Engineering, Mathematics (STEM). Second, it further elaborates whether orthodox women have weaker high-school preparation in mathematics, physics, English and computers. Third, it examines how the establishment of ultra-orthodox women program tailored for their needs fosters female students' success in STEM and decrease dropout from college. Ultra-orthodox students drop-out of school due to limited social support networks, inadequate cultural fit, financial hardship, and lack of perseverance (Meeuwisse, Severiens, and Born, 2010).

Azrieli college of engineering, located at Israel, was selected as subject matter bedrock for investigation because it has both only-women programs and general admission programs open to all students enabling juxtaposition of the programs. The fact it is an an Israeli public institution which trains engineers fits the focus on STEM. It is located in the Ramat Beit HaKerem neighborhood of Jerusalem, between Jerusalem's two major high-tech industrial areas, Har Hotzvim and the Jerusalem Technology Park which are in demand for fresh alumni with engineering 21st century business analytics and big data skills . To fill that need, Azrieli College of Engineering Jerusalem offers undergraduate (Bachelor of Science, B.Sc.) degrees in software engineering, industrial management engineering, electronic engineering, advanced materials engineers, pharmaceutical engineering (cooperation with TEVA), mechanical engineering, as well as a graduate (Master of Science, M.Sc.) in software engineering and entrepreneurship. The college also offers a pre-academic gap-year program to increase diversity by enabling students with weak high-school background to strengthen skills in order to be admitted and succeed in the college's engineering departments. The rest of study is structured as follows: first, a literature review is conducted. Data analysis is conducted using qualitative and quantitative methodologies based on unique data collection on this group of first-generation highly religious minority students which is in itself a major contribution to underrepresented stream of literature (Geiger and Alt, 2014). In so doing, we seek to advance understanding of the choices made by highly religious women who turn a corner in life to pursue higher education. In the process, interesting research questions emerge such as: what are their motivations in pursuing higher education? what challenges do they face? how do they manage them, and what institutional activities smooth their transition to academia?

After data collection procedure, a rigor statistical analysis follows which juxtaposes a veteran ultra-orthodox institution with newly established. Finally, conclusions are drawn with implications for improvement the integration of Ultra-orthodox women in the Israeli higher education landscape.

LITERATURE REVIEW

Previous studies have focused on the impediments that discourage Haredi women from applying to higher education colleges and entering the job market (Golan and Fehl, 2020; Gottlieb, 2007; Hakak, 2004; Malhi, Cohen, & Kaufmann, 2008). The impediments source to the large number of children under their supervision, financial struggle, insufficient governmental sponsorship and scarce work-offers. (Gurovich & Cohen-Kastro, 2004). These factors lead to dropout of Haredi students from college.

Dehan (2013) describes a pioneer Israeli program in discipline of social work tailored for the Haredi population. The preliminary program (1997–1999) was established by the Paul Baerwald School of Social Work of the Hebrew University of Jerusalem (PBS) in collaboration with the Haredi college for women, Neve Yerushalayim College (NYC), and the oversight of the Israeli Ministry of Social Affairs. Its success made the education institutions realize that Haredi community can be semi-integrated into the mainstream education system, a notion which seemed unachievable until its inception because the Haredi sector, request for off-campus, non-mixed gender in the classroom conflicted the principle of non-sectarianism, which was uncompromised value. As a result, the Israeli council of Higher Education started approving similar endeavours mainly in the domains of social work and instruction (Garr & Marans, 2001).

The academic system approach in the last decade has been designed to provide scholarships for minority students in higher education with focus on ethnic, racial, gender and sexual-orientation diversity (Aud, Fox, and Kewal Ramani, 2010;

Fischer, 2007; Hurtado, 2007; Shavit, 2007). Religious diversity on campus was assessed by quantitative proxies such as minority student's GPA (Sherkat, 2007), instead of listening to the students' desperate need for a two-fold supportive infrastructure (academic and emotional) that will help them to sustain strong opposition within their core families for pursuing higher education which sometimes leads to their exclusion from friendly circles in which they were raised from childhood. Consequently, identity and value conflicts among highly religious students were neglected, although psycho-social adjustment stimulates academic success and well-being in college (Tinto, 2012).

Ultra-orthodox students who hail from religious communities, in which higher education is distrusted, are commonly first-generation higher education students, a status demanding a lot of adjustment and creating dual identity (Orbe, 2008). First-generation students who break-away from family traditions can face condemnation from their home communities (Malach, 2014), further complicating their adjustment to college. As simple example, such students are likely to struggle finding a husband after graduation because the way matching criteria within their religious society has been built for generations around stereotypes which rank lowly the people who are attempting to derail from the community path and way of life.

Despite traditional barriers, highly religious individuals are striving to enter institutes of higher education in increasing figures (Malach, 2014; Sherkat, 2007). An exemplar model is the ultra-Orthodox Jews in Israel, who, over the past 12 years, have experienced an 18-fold trend increase of enrolment in higher education, from several hundred individuals to nearly 30% of their adherents (Novis and Rubin, 2019).

Contemporary colleges have been viewed by religious communities as bastions of secularity which derail students away from religious faith towards contradicting scientific evidence about the process universe creation, etc. (Carpenter, 1998). Media commonly portrays colleges unrealistically as vibrant post-secular spaces providing a variety of religious options and undergoing a spiritual renaissance (Astin, Astin, and Lindholm, 2011; Bryant, 2006; 2007; Jacobsen and Jacobsen, 2008; Mayrl and Oeur, 2009; Schmalzbauer, 2013; Schmalzbauer and Mahoney, 2012). In either case, however, ultra-orthodox and secular groups can feel uncomfortable in the presence of one near the other, because their personal lifestyle reflects neither secularity nor spirituality, but a free-choice of behaviour or religious devoutness to conduct their life (Visser-Vogel et al., 2012). A simple example serves the requirement of ultra-orthodox students to dress modestly on campus and not be in contact with opposite gender for mundane matters such as study groups which secular students regard as unacceptable demand. This leads to the need for establishment of an all-women institution for ultra-orthodox students were everybody has the same dress code.

For many ultra-orthodox students who received little or non-exposure to outside world as external media outlets (TV, Radio, etc.) are filtered or entirely blocked by their community since childhood to prevent non-conforming behaviour, enrolling in college signifies a milestone to adulthood, were students are confronted with debates about their real purpose in life (Mubarak, 2007) which has the potential to become a turning point in faith development (Fowler, 2004; Parks, 2000). As an illustrative example, it is common practice for people belonging to ultra-orthodox community to own a Kosher-phone (i.e., cellular device with limited internet access). Their community has separate newspapers in which photos are screened for adherence to modesty rules. The impact of academic knowledge on religious beliefs has been examined in the literatures (Geiger and Alt, 2014; Mayrl and Oeur, 2009). While several scholars found that higher education diminishes religious commitment (Pascarella and Terenzini, 2011), for instance, a change from conservative to liberal views (Reimer, 2010, others postulate contrary findings (Hurtado, 2007). Surprisingly, evidence exists that part of students' population actually undergoes a strengthening of religious convictions to shield themselves (Lee, 2002).

In order to abide to the restrictions, set by ultra-orthodox rabbis, an all-women college institution was established. It is a logical choice because it creates a safe environment for ultra-orthodox students given prejudice posed if they learned as minority status in all open admission university. Bowman and Smedley (2013) discovered evidence that satisfaction from college experience decreases when the social status of the students' religion is low. Religiously committed students tend to suffer from a negative climate on college compared to those who were less religiously strict (Mayhew, Bowman, and Rockenbach, 2014), and incidents of discrimination were commonly reported among fundamentalist group of students (Hyers and Hyers, 2008; Tobin, Weinberg, and Ferer, 2009). This may be particularly so when religious identity is made conspicuous through dress conduct such as garb (e.g. Muslim female students who veil, Cole and Ahmadi, 2003; 2010; Nasir and Al-Amin, 2006; Siddiqi, 2016). Similarly, socializing with students who have similar religious principles was found to be positively correlated with increased religious commitment and overall satisfaction (Small and Bowman, 2011).

Statistical data about ultra-orthodox women enrolment in the Israeli Academia

The ultra-Orthodox are a growing segment of the population in Israel (CBS-Israel, 2011; 2017). They comprise about 11% of the Israeli population and have been labelled 'fundamentalist' because of their religious extremism (Almond, Appleby, and Sivan, 2003; Stadler, 2009). The term "Haredi" refers to multiple Jewish orthodox subgroups that include the Sephardic and Lithuanian Haredi, Hassidim and Chabad Lubavitch (Hacohen & Cahaner, 2011; Levin, 2009, 2011). Serving God and following the Torah's commandments is the ultimate purpose directing Haredi movements (Cahaner, 2011; Cohen 2005; Hakak 2004). Vast majority (over 80%) of Haredi couples marry at a young age and raise large families in accordance with their belief in God's command to multiply (Odenheimer & Ackerman, 2012). Their male and female children are educated from infancy in two different systems.

They typically reside in enclosed neighbourhoods or towns that are entirely shut down on Shabbat (Saturday) and holidays for any sort of work may it be public transportation or merchandise. Ultra-Orthodox Jews cultivate large families (6.5 children on average; Hleihel, 2011). Traditionally, women have focused on the home and family so that their husbands could pursue religious studies. However, because of poverty, women have started pursuing academic degrees with approval community leaders (Rabbis) to generate income for sustenance and livelihood (Baum et al., 2014; Pfefferman and Malchi, 2010).

Approximately 90% of orthodox women after twelfth advance to orthodox seminar, about 41% of them get teaching certificate which emphasizes the lack of STEM education. Orthodox women represent only 5% of studying women (in academic studies) and only 12% of them pursue engineering and sciences studies. As of 2014, among those aged 25-35, 8% of ultra-orthodox (Haredi) women had a college degree, compared to 43% of both secular and moderately religious Jewish Israeli women in that age group. The Haredi community makes up about 11% of Israelis, and many of its male members give up employment and secular higher education in favour of life-dedication to Torah study. The state comptroller report evaluated these programs from 2011 to 2018. Before the government push in 2011, there were some 6,000 ultra-Orthodox college students in the country in a handful of gender-segregated schools tailored to the community. By the 2016-2017 academic year, when some 20 ultra-Orthodox colleges were established, the amount increased to 11,465 students, but fell short of the state's proclaimed target of 14,500. The quantity of students reported includes those enrolled in mandatory pre-college preparatory courses, which can last one to two years before students are admitted to their undergraduate studies. These supplementary education tracks suffer high

dropout rates among ultra-Orthodox men and women in recent years. Among Haredi women, slightly less than half, 47%, who start the journey will complete their education.
The trends mentioned above can be attributed to various factors, including difficulties in bridging educational gaps among Haredi women which although studying STEM subjects through high school, though not at a matriculation level. Other adjustment challenges include the difficulty fitting to academic demands such as testing, and the fact that many of the Haredi students are married and parents of young children upon enrolment. In universities the trend of women dropout continues. Calculus I is a bottleneck in the STEM trajectory.

According to Dr Daphne Getz's research for Samuel Neaman Institute, the antecedent for lack for orthodox women enrollment in college studies is lack of family and community support for ambitious women to fulfill their potential. Israeli higher education system is characterized by a low number of female students and gender scissors in Science, Technology and Mathematics (STEM): Only 25% of students in high-tech related academic fields are women, and only 26% of all employees in the high-tech industry are women. Whether society in general does not value and reward men who take on household chores to enable their wives' personal advancement (including entrepreneurship), it is nearly impossible for a woman to become an entrepreneur at ages 25 to 40 without family support (mainly support from her spouse). Many women find it hard to balance long working hours and long absences from home with raising a family. Also, as women are less represented in this field, they experience feelings of loneliness and discomfort being minority in the classroom.

Another major factor which demotivates women in general and Haredi women in particular is the prospect of find a job after graduation and the commensurate compensation. Recent data by the Israeli Central Bureau of Statistics (CBS, 2011; 2017) shows the average wage gap between men and women stands at 35%. This is an increase of 3% compared with 2015. The average monthly income of a male salaried employee is NIS 11,664, while for women, that average stands at NIS 7,633. Among those with higher education, the gap becomes even wider, reaching 39%. Wage gaps also exist in the hi-tech sector, where young women make 7% less than young men, while women over the age of 40 make 18% less than men. There are high wage gaps in the public sector as well, with the state as an employer creating gender wage gaps.

The 2016 data from the Ministry of Public Security, for example, indicate that the median annual wage for women was NIS 10,000 lower than that of men.

According to data from the Council for Higher Education for the years 2015-2016, 30% of the academic staff in Israel are women, but only 17% of academic professors are women.

Similarly, Israeli women are vastly underrepresented across a number of professional fields, holding roughly a third of senior positions, according to a Women’s International Zionist Organization report released on Wednesday. WIZO’s first-ever Glass Ceiling Index gathered data from government sources, research institutes and academic studies regarding women’s achievements and gender inequalities across a number of fields. Women earn 35% less than men, and among women with higher education, the gap is even larger, at about 39%, the researchers found, based on data for 2017. This also applies to the hi-tech sector, where a young female academic earns on average 7% less than her male counterpart, while women over the age of 40 earn 18% less than their male counterparts. Additionally, only 34.5% of senior executives are women, and only 15.5% of CEOs are women, the report found. The index noted high wage gaps in the public sector. For example, in 2016, there was an average NIS 10,000 annual wage gap between men and women in the Public Security Ministry.

METHODOLOGY

Research design

Data was collected in two complementary methodologies: quantitative surveys of students and exploratory case study based on in-depth interviews with all-women ultra-orthodox college directors. Analysis of this data (Rubin and Novis-Deutsch, 2017) elicited insights.

In the first phase, in order to better understand the arena of an orthodox campus, interviews were conducted with chief directors of Machon Temura, which is affiliated with Azrieli College of Engineering and is associated as its ultra-orthodox branch. In order to triangulate results, similar interview was conducted with the most seasoned ultra-orthodox STEM college in Israel titled Machon Tal. The reason for selecting Machon Tal is that resides in the city of Jerusalem too and offers wide variety STEM degrees unlike most ultra-orthodox colleges in Israel which their offerings are limited to human and social sciences programs. Machon Tal is affiliated with Jerusalem College of Technology. Insightful lessons can be derived from Machon Tal for the benefit of Machon Tmura since its program is 20 years old (unlike Machon Tmura which is relatively new). Due to confidentiality restrictions, it was not possible to administer Google questionnaire to students of Machon Tal, but valuable data was accessible from secondary data sources available online about Machon Tal because of its experience and international recognition. The interview protocol for both directors of Machon Tal and Tmura is described in Table 1. In the interviews, directors were asked to share information about motivations for pursuing these degrees, as well as their concerns, challenges and reflections on the process of building such as unique educational enterprise catered for ultra-orthodox women.

Table 1: Interview Questionnaire used to collect data from Machon Tal and Tmura

1. Who heads the program?
2. What are the requirements to be accepted to the program?
3. In which parts of Israel do students come from? Are there international students?
4. What degree are proposed? Why those?
5. Is there combination of Bible (Midrasha) studies?
6. Where is situated the campus?
7. Is there prerequisite preparatory program (mechina)?
8. Is there mentoring for struggling students (tutoring)?
9. What age groups are students?
10. How does the program support minorities?
11. What's the degree duration?
12. What's the tuition fee? Are scholarships available?
13. How many students enroll each year?
14. What Is the drop-out rate? In which year does it occur?
15. How does the program help the student to find job when they graduate?
16. In which companies alumni work?
17. How long on average after they graduate do students find a job?
18. How do the program support women with babies and pregnant women?
19. Is there dormitory? Kitchen for Kosher food preparation?

The interviews provide the foundation to build a case study approach which is appropriate because of the unique nature of the orthodox colleges. In addition, this approach supports our interest in building theoretical and managerial insights on a topic of Ultra-orthodox college institution that has had little coverage in the literature (Geiger and Alt, 2014; Yin, 2003).

Background on case study data

In 2003, as part of the effort to expand ultra-Orthodox participation in workforce, governmental agencies in Israel enabled higher education accessibility to this population by approving gender-separate academic tracks managed by public universities and private colleges. The ultra-Orthodox population responded positively and between 2004 and 2016, ultra-Orthodox enrolment increased from 600 to 11,000 students (Novis and Rubin, 2019). These remarkable figures were initially offset by the high dropout rate of ultra-Orthodox students: 41%, almost twice the dropout rate of general Israeli society students (Malach, Kahner, and Regev, 2016). Jointly, this data indicates that there is growing interest among ultra-Orthodox individuals to acquire higher education, but that variety of obstacles hinder many of them from accomplishing their mission.

There are several studies on these questions that revealed ultra-Orthodox students pursue higher education mostly for economic sustenance while striving to preserve their conservative worldview (Baum et al., 2014; Kalaagi, 2012), other studies revealed fulfilment and status as antecedents (Dehan, 2013; Laiush, 2014).

Our subject of investigation is ultra-Orthodox women for two major reasons: first, in Israel, ultra-Orthodox college female students currently surpass men at a ratio of 2:1 (Regev, 2016). Second, scarce studies on ultra-orthodox women's educational status in traditional societies (Geiger and Alt, 2014) indicate that women are agents of societal change (Herzog and Braude, 2009) that have the power to revolutionize their community (Thomas, 2001). Acquiring higher education can serve as an engine driving such shift in traditional mode of life and structure of family cell breaking their community leaders' control over information flow and decision-making (Ammerman, 2013; Laiush, 2014; Sivan and Caplan, 2003).

Case 1: background information about Machon Tmura

Azrieli college of engineering in Jerusalem offers orthodox women the opportunity to study in Seminar (Machon in Hebrew language) Tmura. In Seminar Tmura, orthodox women can choose between industrial engineering or programming engineering degree. They learn in separate campus with women only. Tmura opened as the result of the growing demand for adapted study environment for ultra-orthodox women. Ultra- orthodox women were already integrated in low tech domains such as quality assurance, Seminar tmura opened the door to advanced high-tech disciplines. As part of their integration into the industry, the students receive professional and practical workshops and personal counseling in order to prepare them to the industrial arena and to deal with the wide array of jobs. Among the companies that employ the graduates of the Tmura seminar are prestige Israeli conglomerates such as: Intel, El Al, Mobileye, Cisco, Israel Aviation Industry, etc. The graduates earn high-paid salaries that fit the high-tech world in combination with a work environment suitable for ultra-orthodox women.

The curriculum is completely adapted to the values of the seminar and includes the guidance of an educator, religious studies, Halacha (Jewish faith) and lessons as a solid foundation and leads to the establishment of a home of Torah (Bible), along with a bachelor's degree in engineering. Tmura Seminar and Azrieli College work together to build a network which integrates the women in the prestigious high-tech industries through collaborations and connections that have also been created by the alumni and their good names that goes before them.

In 2018, the seminar inaugurated a new campus with renovated class rooms and state of the art computer labs. In the campus, more than 100 young women are studying. Tmura exists since 2011 but till 2018 the students had one major to specialize: software engineering. From 2018 hence, they can also choose to learn industrial engineering with information system specialty.

According to Dini Vice, the seminar's director: "*The Tmura Seminar was established to provide the best possible answer, in every way, to the huge demand* *for engineering studies with the spiritual and professional level. We are guided by an uncompromising vision: to give every girl, from everywhere, the great opportunity to integrate into a dignified and rewarding employment by studying for a sought-after and prestigious degree - while maintaining Jewish principles and values."*

Alumni graduates are employed in the Department of Aircraft Control Systems at El Al Airlines, in the development departments at Intel, such as the Department for the Creation of Devices with Visual Capability at Intel. At Mobileye - in the development of technology for autonomous vehicles. Others work for Orcam, Cisco and other big companies.

Case 2: background information about Machon Tal

Established in 1990, Machon Tal pioneered Israel's women-only Engineering and Technology College. Tal's high-level academics combined with a comprehensive Jewish studies program, have earned the College an international recognition attracting Jewish students from diaspora. The campus serves over 2,000 women studying in various undergraduate and graduate degree programs including Industrial Engineering, Computer Sciences, Bio-Informatics, Accounting, Business Management, and Nursing.

Strategically located in the Givat Shaul neighborhood of Jerusalem, Tal Campus spreads over 5,100 sq. m of rented space. In addition to classrooms, meeting rooms, a library, high-tech laboratories and simulation labs, Tal offers comfortable dormitories and special services for young mothers such as on-site day-care.

Alumni have a job placement rate of about 90% and are sought after candidates in Israel's rapidly expanding high-tech industry. Many have attained top positions in leading firms and start-ups in industry such as Intel, Texas Instruments, Check Point and IBM, among others, and are involved in classified R & D projects for Israel's defence industry. Others choose to pursue academic careers and are instructors at top niche colleges and universities.

Tal's Nursing program is well esteemed as one of the top nursing programs in the country. Students continuously attain among the highest national averages on national board examinations and have a high pass rate. Graduates are highly sought-after candidates for positions throughout Israel's healthcare system.

The majority of students in JCT's pre-academic program belong to Haredi and Ethiopian communities. Many are married with kinds to support and encounter financial pressure putting them at high-risk of dropping out.

JCT's Mechina (pre-academic preparatory program) gives an opportunity for students who lack the necessary academic background to gain high school equivalency and become college-ready within 16 months. Students begin with a three-month pre-Mechina program in mathematics and English to assess their ability to study in a classroom environment and their academic aptitude. Upon successful completion of the Pre-Mechina Program, students enter a full year of Mechina which includes intensive lectures in math, English and physics in preparation for their academic studies. JCT's Mechina Program builds a springboard that gives a chance for underprivileged populations to encroach academia.

The College prioritizes scholarships as well as living stipends to students in need. JCT offers ongoing support throughout their degree, including tutoring, counselling, pre-exam preparation courses, CV writing and presentation workshops, an annual job placement conference, and guidance from the College's job placement staff. The system offers equal opportunities to succeed for students with learning impediments in order to enable a fair chance to thrive and unlock their potential.

In October 2019, JCT established Israel's first Computer Science Degree in English for women. The dual curriculum encompasses academic content coupled with bible study. Stipends are available for students in need, especially immigrants. Interestingly, the Ethiopian and Haredi women in Israel share similar obstacles in terms of pre-designated role within family to take care of kids and old-fashion tradition that are difficult to break from. JCT identified this opportunity to expand their admission pool and was one of the first Israeli institutions of higher education to offer tailored academic degrees for the Ethiopian women too in the fields of STEM. Promoting Ethiopian women enhances their quality of life within the Ethiopian community and contributes for diversification of the broader Israeli society.

DATA ANALYSIS

A questionnaire was sent to Machon Tmura students in the track of industrial engineering. It is important to note that Machon Tmura admits only ultra-orthodox Jewish women students (Men students are not enrolled in Machon Tmura). Similar questionnaire was administered to the secular Azrieli general college targeting non-orthodox female students from departments of industrial engineering and software engineering. The general college of Azrieli admits men too, but they didn't receive the questionnaire. The questionnaire was composed by numerous independent faculty members to ensure content validity. It was administered by filling Google website prepared by an undergraduate student. To encourage students to fill questionnaire, follow-up email reminders were sent to students by Head of machon Tmura and senior instructors in Azrieli college. The sample was composed of 143 students: 70 from Azrieli, 73 from Tmura. The response rate was approximately 90% in Machon Tmura and 75% in Azrieli. Table 2 describes the questionnaire items.

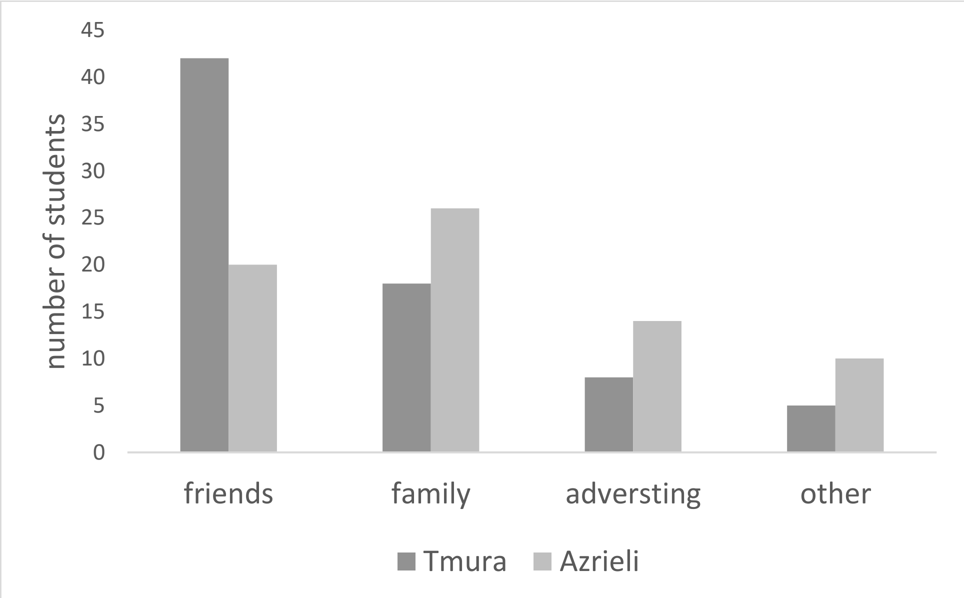
Table 2: Google Questionnaire Administered to Students of Ultra-orthodox Machon Tmura and Secular college of Azrieli

1. In which campus do you study? (Azrieli-Bet hakerem or Tmura-Har HaZofim)
2. What is the main source you heard about the college form? (Friend, family, advertising, other)
3. What influenced you to study engineering in this college? (livelihood, career, satisfaction feeling, interesting studies, to know interesting people, to pass the time)
4. In which school year are you? (mechina, 1^{st}, 2^{nd}, 3^{rd}, 4^{th})
5. Do college lecturers (women) serve as a role model for you?
6. Do you know female students that drop of studies and why?
7. Do you work in parallel with your studies?
8. Do you work in a field related to your studies?
9. Please rate the importance of the following topics in your life: marriage and family, career, friends, spirituality, materialism.
10. How much the studies are difficult for you?
11. Did you receive assistance in those fields: mehina, scholarships, tutoring, help in finding job?
12. Did you learn those fields in high school: mathematics, physics, English, Computers?
13. Did someone in your family learn engineering?
14. Do you think now you will apply for master degree after graduating?
15. How do you see your professional future?

The answers were analyzed by SPPS software, using chi-square test to juxtapose between the two groups: Azrieli and Tmura students, ultra-orthodox vs.non-ultra-orthodox students (N=143 (70 non-ultra-orthodox at Azrieli; 73 ultra-orthodox at tmura)). Next, the answers for each question are described:

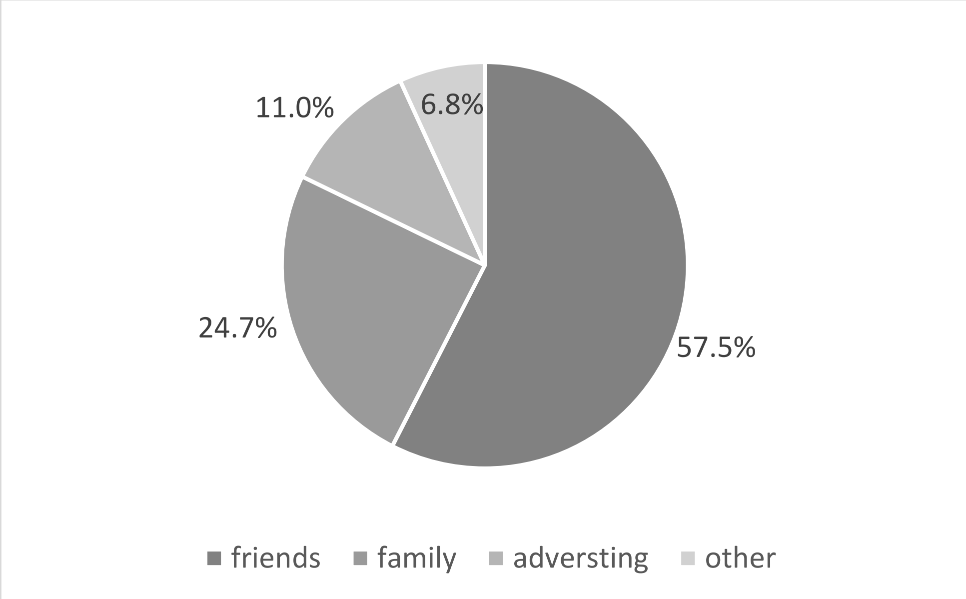
First, what is the main source you heard about the college from (Friend, family, advertising, other)? The statistical analysis indicates a clear difference between the two groups in the category *friend* (Figure 1.a). As a matter of fact, in the category of friends as source for information about college there is ratio 2:1 in favor of Tmur, whereas in ratio of public Advertisement as source of information about college the outcome is vice versa 1:2 in favor of Azrieli. It can be explained by the logic that orthodox society is blocking information from outside sources to reduce potential stream of non-orthodox content or influence on the community members, therefore, remaining source of information becomes friends. In contrast, in the secular society the main source of information is public media advertisement. The Chi-square test shows a dependency between college type and source of information (chi square value=12.5 with significant p-value=0.005). When conducting a pairwise chi-square test between college type and friends the statistical difference is evident (chi square value=12.21 with significant p-value=0.00)

Figure 1.a



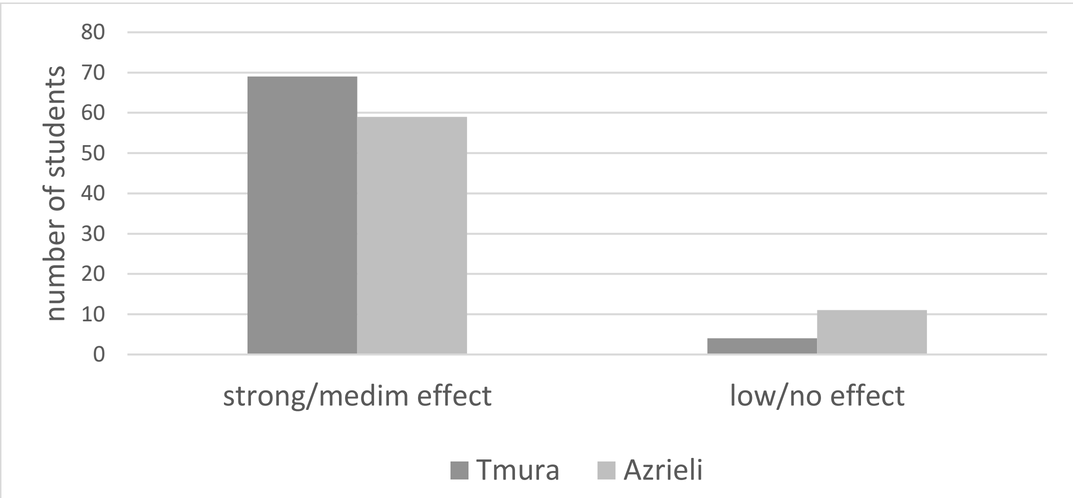
In order to fine-grain what is the distribution of source for information within Tmura college, Figure 1.b is portrayed. It elucidates a significance difference among sources of information for prospective students about Tmura within ultraorthodox community (chi-sqaure=46.288, p-value=0), where *friends* account for 57.5% of the sources while advertisement accounts for merely 11%.

Figure 1.b



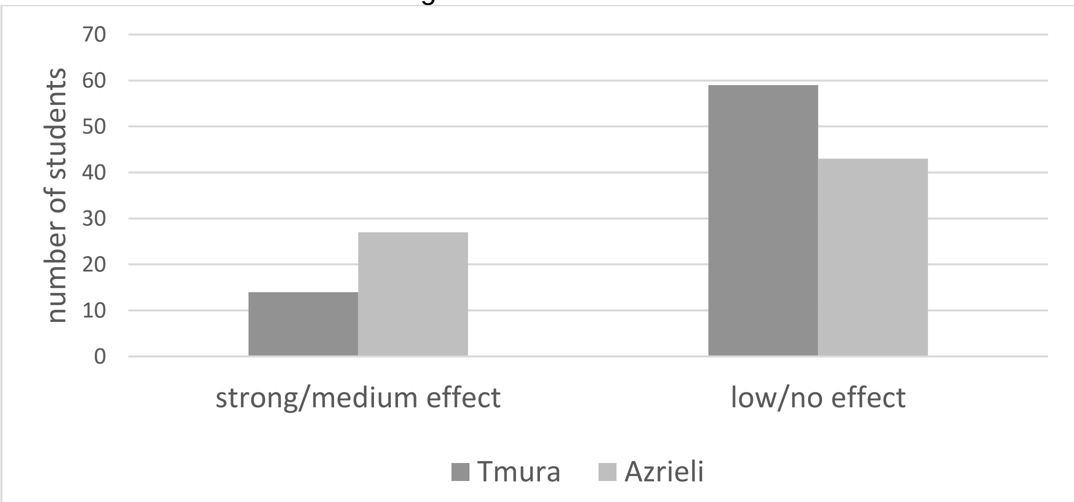
Second, *what influenced you to study engineering in this college?* There is a clear difference between the two colleges in the impact of *livelihood* as a motivating factor to enroll in college (Figure 2.a). For most of the students in Tmura (69/73) livelihood is important. The ultra-orthodox women enroll in studies in order to learn a profession that can help put food on the table for their families in need of income since they are the main providers of sustenance within family (husbands devote life for bible study). The Chi-square test shows a dependency between college type and livelihood as motivator (chi square value=3.98 with significant p-value=0.04).

Figure 2.a



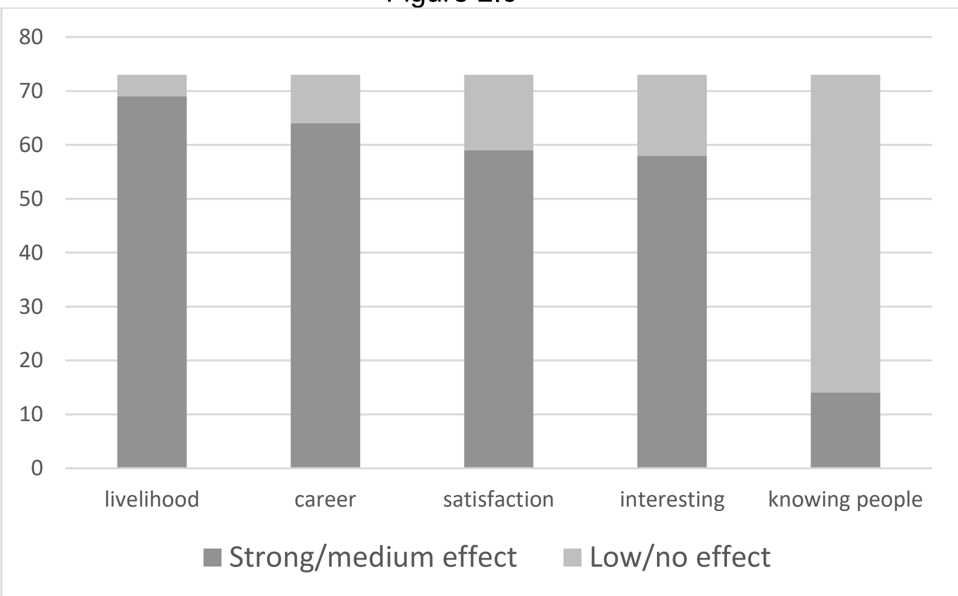
Interestingly, there is a clear difference between the two colleges in the category of *getting to know* people as a motivating factor to enroll in college (Figure 2.b). The ratio in Azrieli compared to Tmura is 2:1 in favor of Azrieli students eager to meet colleagues on campus. In contrast, the Ultraorthodox community is tight in nature where new acquaintances are created mainly through family network. Furthermore, meeting people outside the family circle is regarded as religiously inappropriate. The Chi-square test shows a dependency between college type and *getting to know people* as motivator (chi square value=6.57 with significant p-value=0.01).

Figure 2.b



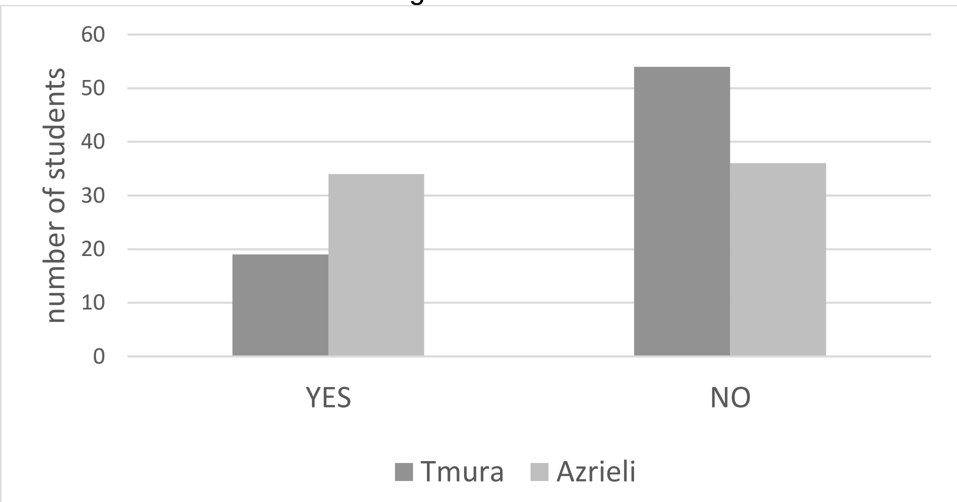
In order to fine-grain what is the distribution of motivating factors to enrol in college within Tmura, Figure 2.c is portrayed. It elucidates insights consistent with theory showing that 94% of students within Tmura are motivating by livelihood, 87% by career, and 19% by opportunity to know people.

Figure 2.c



Next, research question investigates *do college lecturers (women) serve as a role model for you?* There is a clear difference between the two colleges in role female instructors play as role model. While in Azrieli 48% of students highly regard the lecturer as a role model, in Tmura this factor reduces to 24% (Figure 3). In Tmura the lecturers serve less as a role model for the students because in the ultraorthodox community the Rabbi (male Jewish sage) is the address for guidance in any matter. It is important to realize the Rabbi leadership position extends far beyond Biblical rules to daily decision making. For example, the orthodox political parties adhere to the Rabbi's authority too. In contrast, in the secular community at Israel women are more feminist in character therefor tend to admire female instructor who managed to reach high echelon position as professor by acquiring a Ph.D. degree. The Chi-square test shows a dependency between college type and instructor capability to become role model in eyes of students (chi square value=7.78 with significant p-value=0.005).

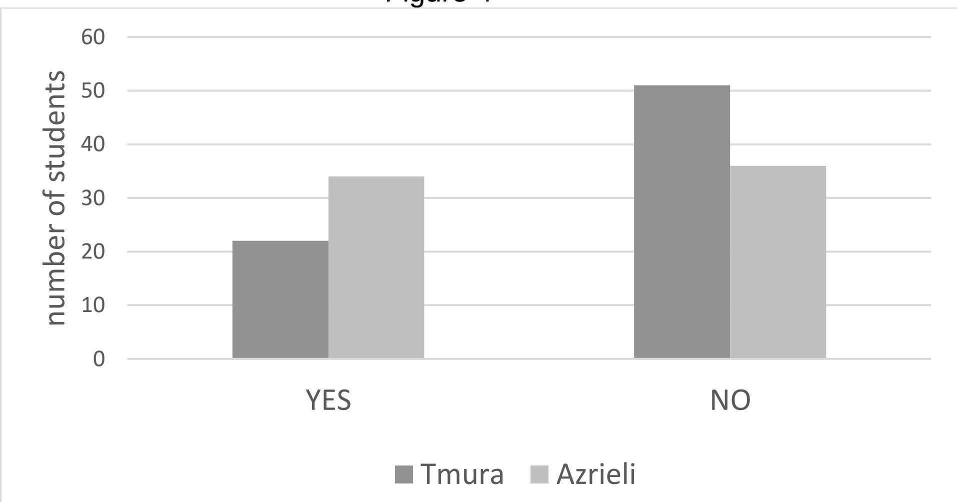
Figure 3



Next, research question investigates whether student *works* simultaneously *with their engagement in studies?*

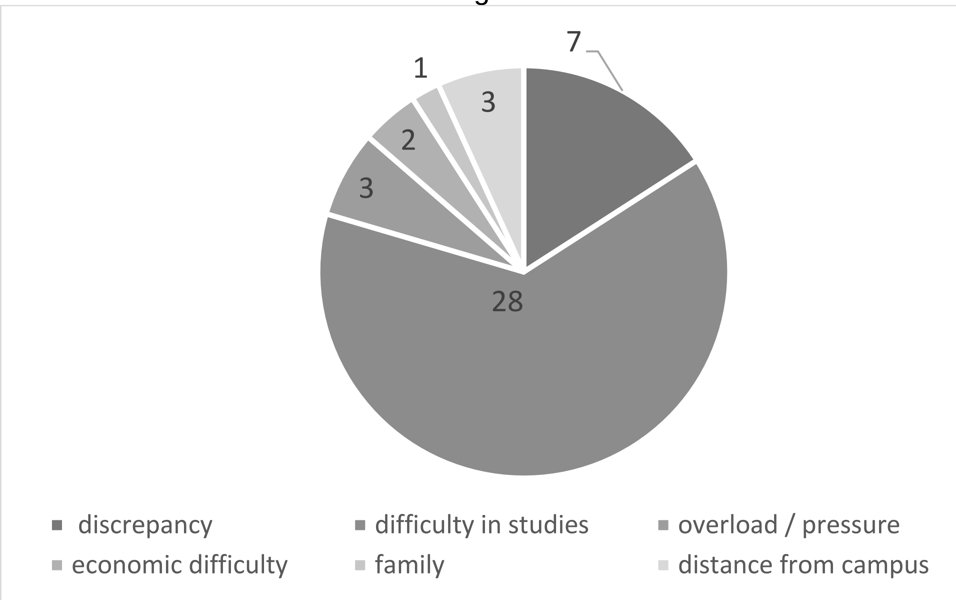
There is a clear difference between the two colleges employment statistics. While in Azrieli 49% of students work during the period of college, in Tmura this factor reduces to 30% (Figure 4). Several factors contribute to this discrepancy. For example, In Tmura most ultraorthodox students got married in their early 20's and have at least two or three kids which makes work in addition to study an overload. Also, the ultraorthodox community limits the ability to work in student's related jobs from outset due to religious restrictions. Simply, many jobs that are regarded non-appropriate in terms of modesty (mixed environment between men and women). Consequently, students in Tmura college are in dire need for scholarship to fund their studies. The Chi-square test shows a dependency between college type and employment (chi square value=21.12 with significant p-value=0.00).

Figure 4



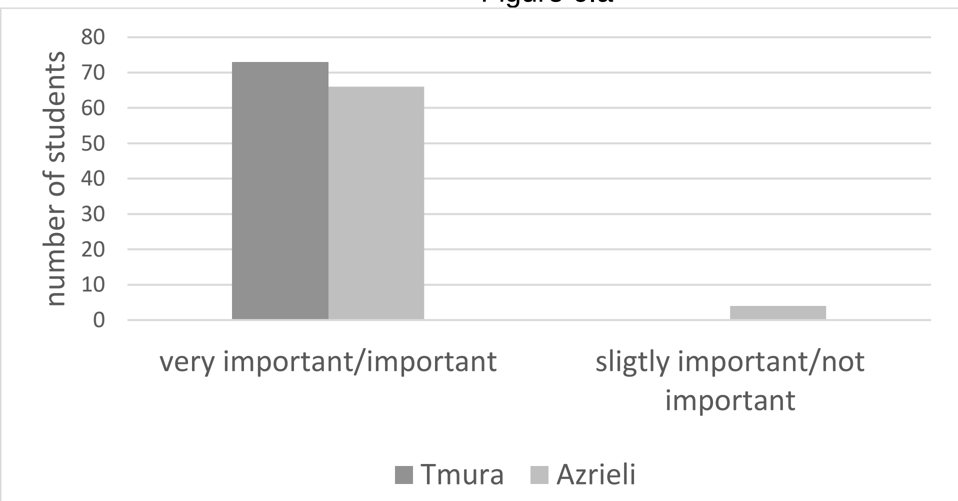
Since the rate of drop-out differs sharply between college (Azrieli=10%., Tmura=1/40), the students in Tmura were asked what was in their opinion antecedents for dropout from college. But beforehand it should be realized that although the dropout from Tmura college is low, the rate if dropout from Tmura preparatory mehina (pre-college first year program) is 10%. In order to fine-grain the antecedents, Figure 5 is portrayed. It elucidates a significance difference among sources of dropout within an ultraorthodox community (chi-sqaure=72.72, p-value=0), where *difficulties in studies* account for 64% of the causes effect for dropout. This phenomenon stems from the poor preparation with which ultraorthodox students arrive to college and will be the subject of next research questions.

Figure 5



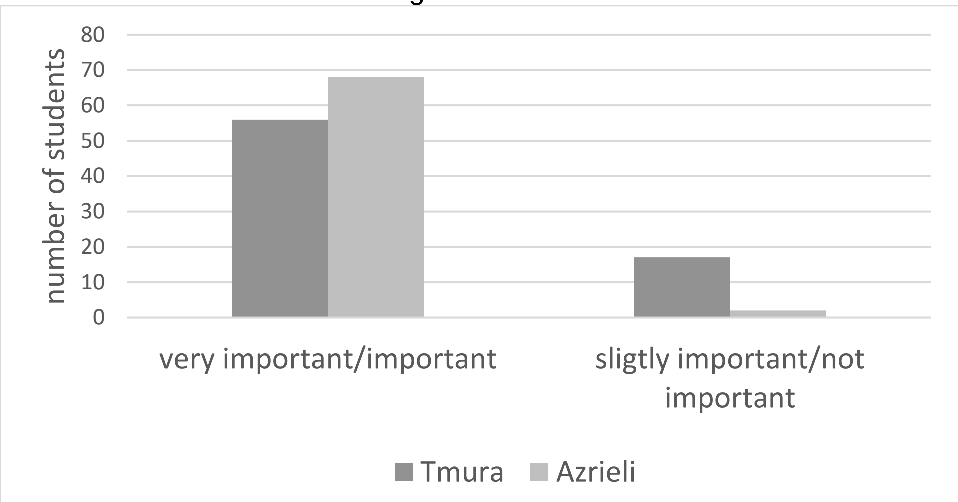
An important aspect which impacts dropout from college is the priorities of students in life. In order to comprehend this important aspect, we asked students *Please rate the importance of the following topics in your life: career, family and spirituality.* There is a clear difference between the two colleges in the importance of family in the eyes of students. In Tmura this factor is 100% important while in Azrieli it is 94%. (Figure 6.a). Several factors contribute to this discrepancy. In the ultraorthodox community, establishing a family is regarded as highest priority in life. People who are not married are regarded as unsuccessful accomplishing life's purpose. The Chi-square test shows a dependency between college type and family value (chi square value=4.29 with significant p-value=0.04).

Figure 6.a



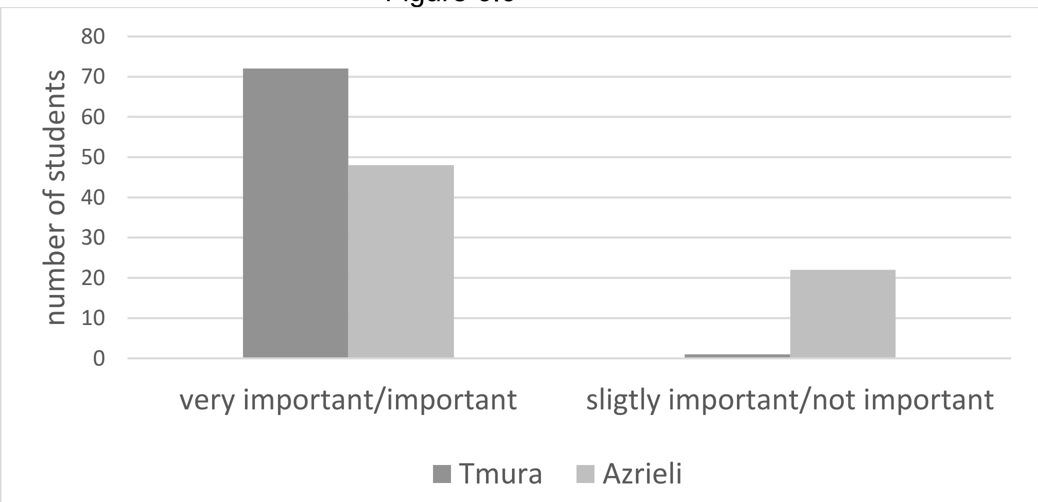
Similarly, we asked students to rate the importance of *career* in life. There is a clear difference between the two colleges in the importance of career in the eyes of students. In Tmura this factor is 76% important while in Azrieli it is 97%. (Figure 6.b). Several factors contribute to this discrepancy. In the ultraorthodox community, career is traditionally less relevant compared to establishing family and educating kids to follow bible rules. In contrast, the secular community encourages women to develop a successful career as part of a feminist perspective about life. Furthermore, women are encouraged to take over jobs that in past were traditionally occupied in secular society by men. The Chi-square test shows a dependency between college type and career value (chi square value=12.94 with significant p-value=0.00).

Figure 6.b



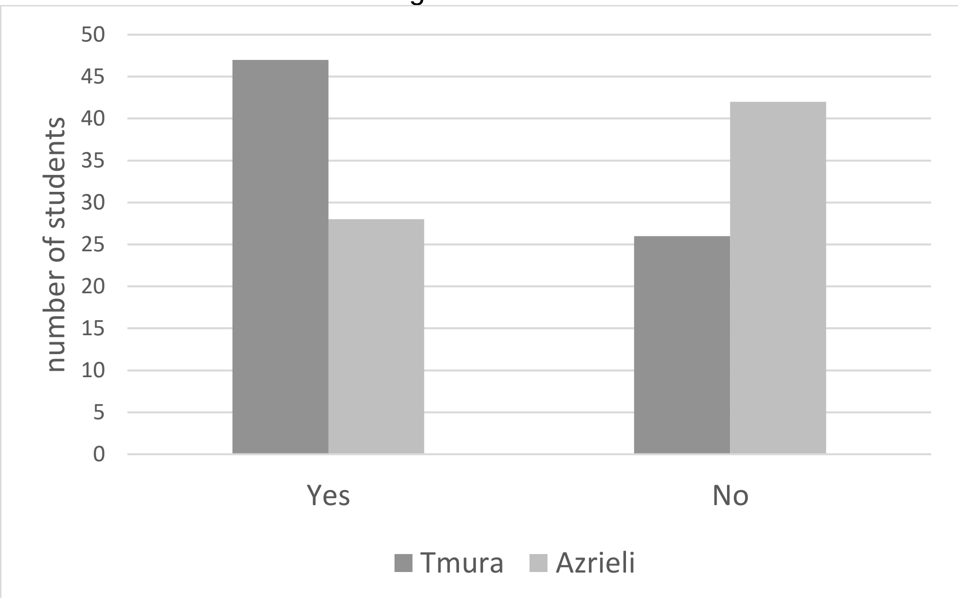
Finally, we asked students to rate the importance of *spirituality* in life. There is a clear difference between the two colleges in the importance of career in the eyes of students. In Tmura this factor is 98% important while in Azrieli it is 68%. (Figure 6.c). Several factors contribute to this discrepancy. In the ultraorthodox community, spirituality in installed from childhood in the DNA of kids through rigor bibles studies in Ulpana (women ultraorthodox school). The knowledge of bible rules and commentaries is regarded as esteemed value in the community which improves chances of finding a prestige companion for arranged marriage. In the secular community, the spirituality is studied few hours if at all in school programs and exams on its content are not rigor. The Chi-square test shows a dependency between college type and career value (chi square value=23.92 with significant p-value=0.00).

Figure 6.c



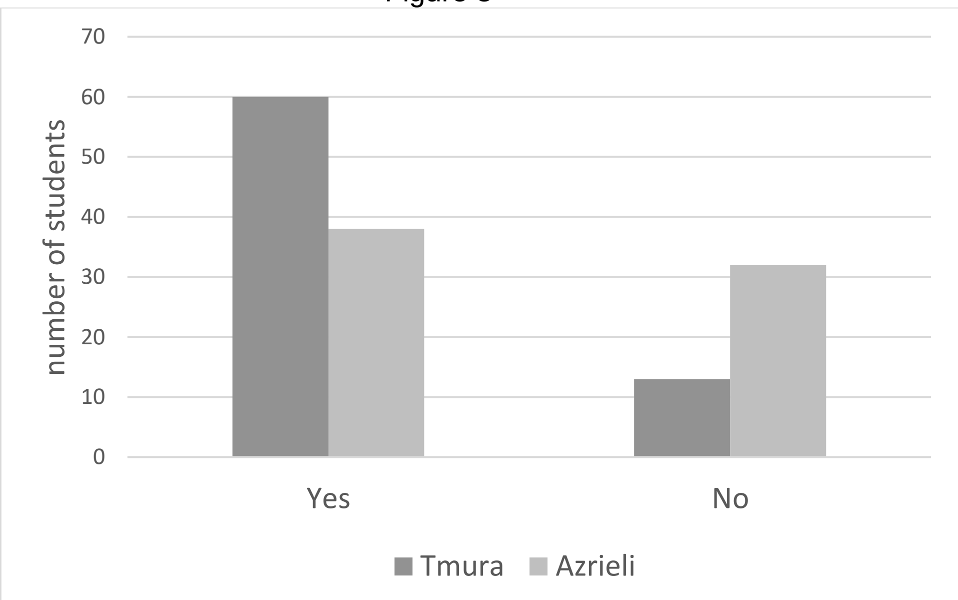
Next, we inquired whether students to enrolled in a preparatory program (Mehina) administered by college in advance of beginning formal degree classes in order to strengthen their skills. There is a clear difference between the two colleges in this category. In Tmura 64% of students enrolled in Mehina, whereas in Azrieli the level of enrollment was 40%. (Figure 7). Several factors contribute to this discrepancy. In the ultraorthodox community, the education is geared towards learning Bible related topics. Consequently, the students admitted to college are beginning from low starting point and need reinforcement before entering the classroom degree courses. The Chi-square test result supports a dependency between college type and mehina enrollment (chi square value=8.52 with significant p-value=0.00).

Figure 7



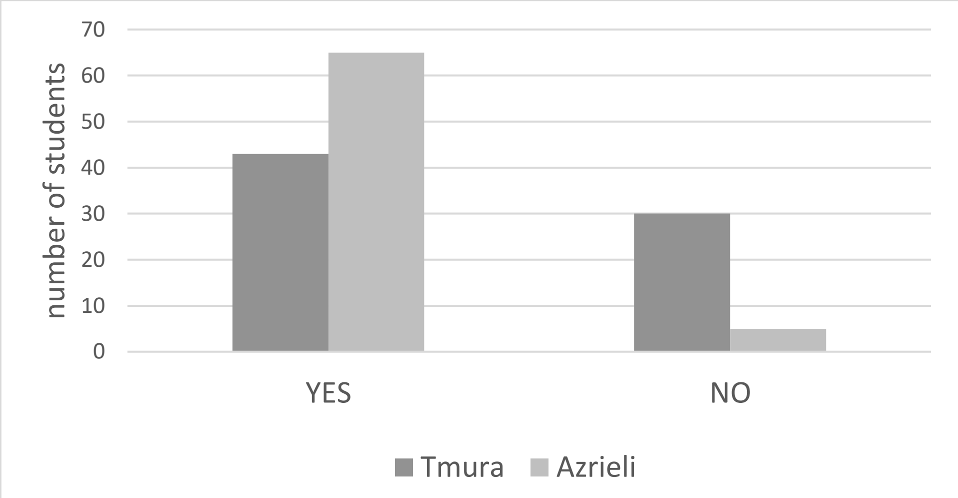
Next, we inquired whether students to receive scholarships in both college. There is a clear difference between the two colleges in this category. In Tmura 82% of students receive scholarships, whereas in Azrieli the level of scholarship acceptance was 52%. (Figure 8). The ultraorthodox students need scholarships because neither them nor their spouses work, making the scholarship a survival necessity for the household. The Chi-square test result supports a dependency between college type and scholarship acceptance (chi square value=8.52 with significant p-value=0.00).

Figure 8



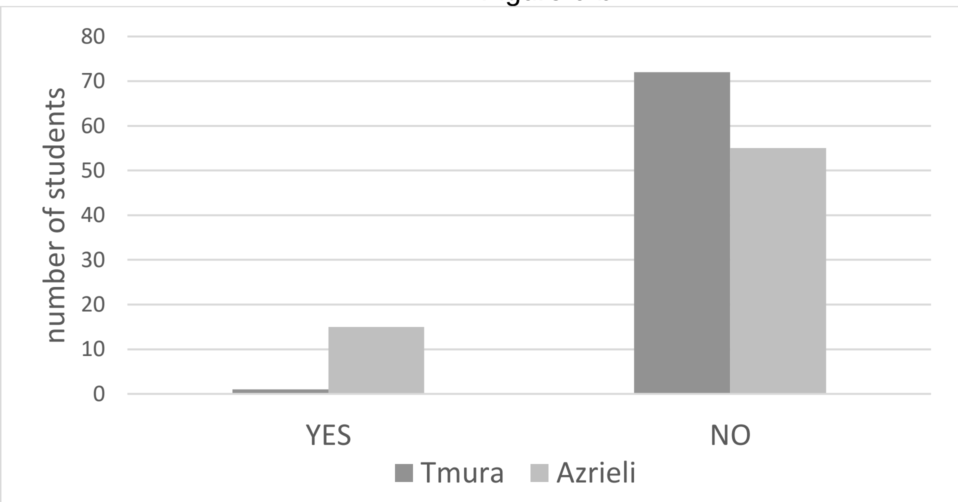
In order to delineate the exact level of students coming to both colleges, we asked students *Did you learn those fields in high school towards a certificate exam (tehudat bagrut): mathematics, physics, English, and Computers?* There is a clear difference between the two colleges in mathematics. In Tmura this figure is 59% while in Azrieli it is 92%. (Figure 9.a). in the secular community, mathematics is highly regarded as a skill necessary to succeed in life, whereas in the ultraorthodox community is a tool for financial transactions. The Chi-square test shows a dependency between college type and mathematical training (chi square value=22.28 with significant p-value=0.00).

Figure 9.a



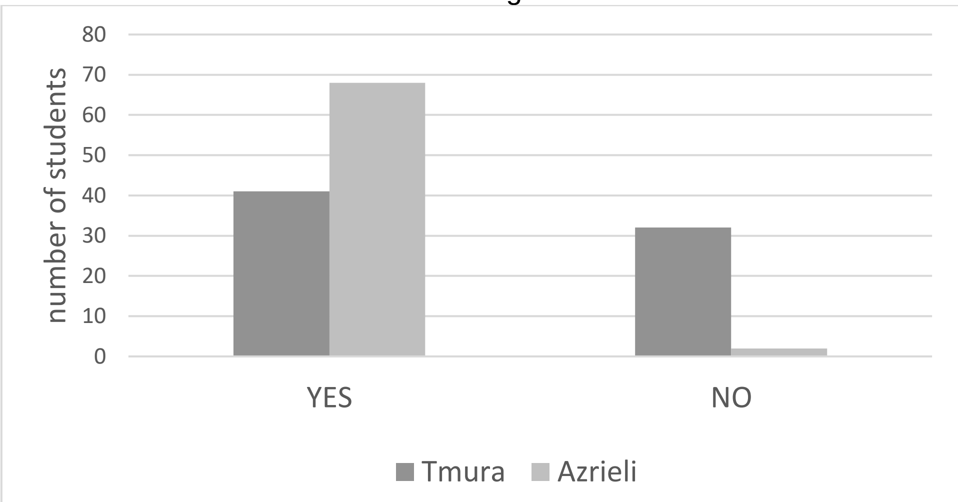
Similarly, there is a clear difference between the two colleges in physics. In Tmura this figure is 1% while in Azrieli it is 27%. (Figure 9.b). In the secular community, physics is highly regarded as a way to explain the creation of universe, whereas in the ultraorthodox community physics is an irrelevant topic as the universe was regarded as created by God and any contradictory claim is a sinful thought. The Chi-square test shows a dependency between college type and physics training (chi square value=14.47 with significant p-value=0.00).

Figure 9.b



Finally, there is a clear difference between the two colleges in English too. In Tmura this figure is 56% while in Azrieli it is 97%. (Figure 9.c). In the secular community, English is highly regarded as a way to communicate with the global world, whereas in the ultraorthodox community English is associated with a foreign culture which may expose students to non-Jewish content. The Chi-square test shows a dependency between college type and English studies (chi square value=33.11 with significant p-value=0.00).

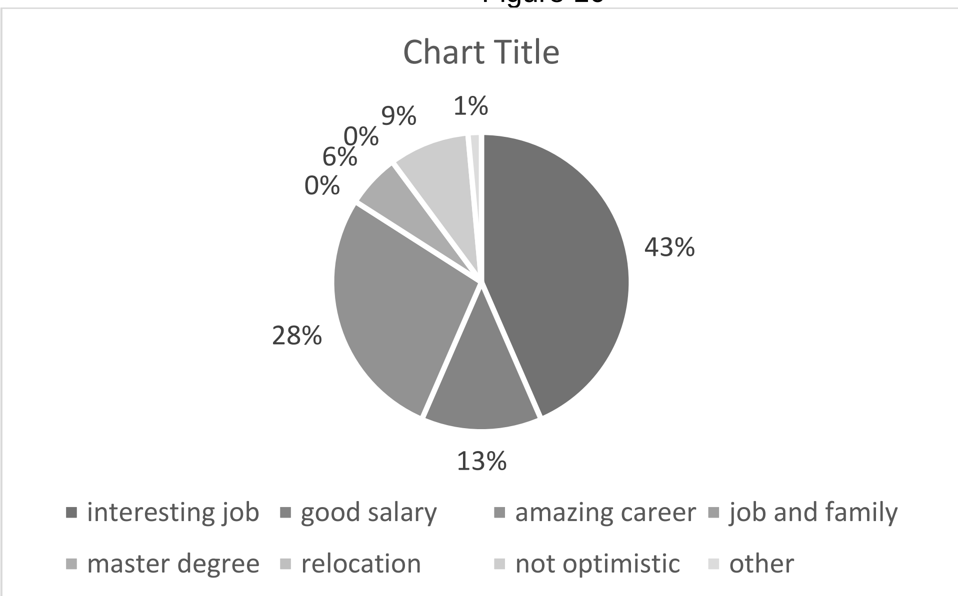
Figure 9.c



In computer science there was no significant difference between colleges, because the ultraorthodox community realizes that information systems and knowledge about computers is essential skill in the 21st century. Furthermore, the ultraorthodox rabbis have instructed their community members to install software blockers (labeled as Kosher software devices) that ensures internet access isn't being used for non-moral purposes (Golan and Campbell, 2015).

Finally, students within Tmura college were inquired about their *future plans* after graduation. Figure 10 describes the distribution of answers. Interesting jobs (43%) and successful career (28%), and good salary (13%) were dominant replies. This provides evidence that Tmura students prime goal is to seek way to provide sustenance for their families.

Figure 10



DISCUSSION

Interview with Dini Weiss, Director of Tmura seminar and Rahel Rozenberg, counselor at Tmura seminar

Dini Weiss, Director of Tmura, introduced the seminar by explaining about the opportunity Tmura gives for orthodox women to learn engineering in high level. Indeed, graduates finish the program with engineering degree from Azrieli college of engineering which is widely recognized in Jerusalem and across Israel. The admission requirements to be accepted into the program are the same as Azrieli college so most of the women need to learn in preparatory program (mechina) or preparatory courses to compensate for the lack in scientific subjects. They usually take soft exams during this preparatory program that lasts half a year (40 hours a week).

About 10% of the women drop out after mechina, according to Dini Weiss most of dropout were originally accepted under probation. After completing the mechina they can choose between two tracks: industrial engineering and software engineering. There are about 20 students in each degree every year. While asked why there are only 2 options for degree Mrs. Weiss explained that other degrees in Azrieli college of engineering are masculine degrees and those two tracks are acceptable for women as it's common in ultra-orthodox community.

The campus is located in Har HaZofim neighborhood in Jerusalem where there are settled most of the high-tech enterprises in Jerusalem. In Beit Adar campus women learn in renewed classrooms, practice for computer courses in well-equipped computer labs and they also have their own kitchen that can be used all day long to cook Kosher food. There is no dormitory on campus.

Beyond the academic courses that are provided by the Azrieli college esteemed teachers, women in Tmura seminar enjoy an extended enrichment content such as religious learning, Tours all over Israel and other extracurricular activities. Mrs. Rozenfeld quoted one of the lecturers that describes the students as a group in opposition to individual students that he usually teaches in other colleges, and she postulated it's probably the most important factor contributing to the women success in their study. Mrs Weiss added "*They are here on for each other and no one is left beyond*."

The second factor that came out from the interview is the shell of educational support they receive from the staff such as help in every problem encountered during the studies (the staff directly speaks with the teachers to help solve problems). For example, there are private teachers for women that gave birth and have new baby, mentoring and more.

To help graduates find a job, the seminar organizes several workshops such as "*How to write curriculum vitae*" or "*Work Interview preparation*". Another useful resource is cooperation with start-ups through internships. According to Mrs Weiss graduates find jobs on average nine months after graduation.

Ms. Weiss provided additional statistical figures. For example, 80% of women just finished high school, 15% are in professional reconversion and the last 5% are women who start learning later. In terms of location, about 60% are from Jerusalem, 25% from cities around Jerusalem such as Bet Shemesh or Beitar and 15% come from the north, south and center of Israel (because Tmura has competitors in these regions).

The program length of studies extends about 4-5 years (including the half year of preparatory program), composed from 2.5 years in regular morning studies and 1.5 years in evening studies. It gives the opportunity to the students to start working during their studies, indeed, about 50% of them do so. According to Mrs Rozenberg' Machon Tmura gives a chance to the students to start building their resume already while studying and helps them to finance their studies. She also explained most of the girls receive scholarship from "Keren Kemach" for ultra-orthodox students (about 70% of tuition costs), and more scholarships such as subsistence scholarships are

available to apply. The tuition cost of seminar Tmura is equivalent to Israel public university tuition costs based on council of higher education framework.
Alumni integrate into prestige high-tech companies such as Intel, Mobileye (about 30 alumni are working there today), Cinamedia. Another major place for employment is the government offices, since Jerusalem is the capital of Israel where most ministry offices are based.

Interview with Director of Machon Tal

In Machon Tal, about 2000 women are studying overall. They come from all over the country, most of them just after matriculation (bagrut - high school diploma). The conditions to be accepted are a combination between psychometry (SAT Israeli exam), with a threshold grade of over 550 cumulative score, Till exam, bagrut grades in mathematics and English. Candidates with lower grades than required or without bagrut certification can learn in preparatory program (mechina in Hebrew language) from 3 to 12 months. Each candidate also needs to undergo and pass a rigor personal interview.
The Director described the diverse possibility Machon Tal offers with eight B.A and B.S.C degrees in fields such as electro optics engineering, nursing, industrial engineering, accounting, business management and 5 Master's degree such as bioinformatics, electro-optics or nursing. Mahon Tal offers also an English program in computer sciences.

As is accustomed in the orthodox community, students come to know about Machon Tal, from word-of-mouth through network of friends or family. Accoring to director: "*What's important for the candidates is to know the college will respect their religious values while giving them rigor education studies.*" Students are offered the possibility to study with religious women group only (Midrasha in Hebrew), and they learn Bible studies. Importantly, students can live in dormitory which is adapted to their Kosher needs. All conditions required to study and succeed while keeping their religious observatory values and needs.

Students who need reinforcement and support can receive tutoring lessons. Young mothers can enjoy a nursery service and breastfeeding room. Tuition costs in Mahon Tal are equivalent to university tuition according to Israeli Council of Higher Education fee and students can apply to several scholarships (excellence scholarships, scholarships against voluntary hours or Jewish studies hours) and loans. Students are prepared for their entry into industrial work environment with demo interview preparation sessions, curriculum vitae writing workshops, individual counseling discourse and job counseling opportunities meant to smooth transition.

Methodological limitations

Several methodological limitations should be acknowledged when evaluating this study outcomes. First, the study collected data in Israel from two colleges that are both located in the city of Jerusalem. Ultra-orthodox population of students isn’t as homogeneous as assumed (Novis and Rubin, 2019). The ultra-orthodox community in Israel is diversified in its strictness of following biblical commands, and code modesty dress. Religious enclave communities have continually been influenced by the Western modern and post-modern values they reject (Eisenstadt, 2000). Although the boundaries are blurry, the difference between worldview can be stark. Furthermore, some ultra-orthodox communities are more involved in the country secular life. For instance, members of the ultra-orthodox community located in Judea and Samaria parts of Israel are proudly serving in the Israeli military, some of them reaching higher echelon ranks, whereas the members of community in Jerusalem tend to be exempted in order to entirely devote themselves to traditional Bible studies life style been kept strictly for generations. Thus, a sample composed of colleges in the west bank of Israel could have yield different statistical results. Obviously, communities were the level of abidance to religious commands is lower would naturally integrate in a smooth manner to college environment after being exposed to secular life style during their three years of army service. Similarly, the Dati-Leumi (patriotic-orthodox) community in Israel who studies in women colleges located in the centre of Israel such as Kiryat Ono academic college and the Sami Shamoon College of Engineering at the Southern city of Ashdod are composed of ultraorthodox students with a different level of adherence and interpretation of biblical obligations than students admitted to Jerusalem colleges. Since the ultra-orthodox community isn't homogenous, there is a need to adopt an approach of multiple traditionalisms (Kalaagi, 2012). Geiger and Alt (2014) followed this path in the pursuit to comprehend the journey to higher education of female students who belong to the Jewish unique Chabad community and discovered a radical shift within the Haredi community that prompted an ideological transformation legitimizing Haredi women's enrollment in non-Haredi institutions. The commitment and support of parents, in-laws and spouses, who shared in childcare and household chores relieved the Chabad Haredi students load and unlocked their potential for academic success.

Second, the study focused on STEM degrees. Most of the ultra-orthodox community who does pursue higher academic education are studying social sciences with emphasis on becoming educators in religious teaching institutions. Third, utilizing non-ultra-Orthodox interviewers to collect data may have caused informants out of loyalty feelings to portray their home-community in a way which is not entirely accurate in order to preserve its dignity, save their personal face, or avoid embarrassment from their family, relatives, and friends.

Integrating ultra-orthodox groups, women in particular, into higher education system should be of interest to political policymakers. This group is under-represented in academic institutions worldwide (Novis et al, 2016). Consequently, women end up being hired in non-professional, lower income jobs, becoming unfulfilled potential for their country's economic growth (UNDP [87]). It is a pity because higher education affords a golden opportunity for interaction between ultra-orthodox enclave communities and the secular society surrounding them, leading to mutual understanding and open dialogue. This can lead to bridging gaps by prejudice reduction on both sides and to the emergence of common grounds, critical to build a strong Israeli society (Bowman and Brandenberger, 2012).

The importance of creating job opportunities for ultra-orthodox women stems from a recent report of Israel Democracy Institute (2020) which finds that household income for ultra-Orthodox families is 58% lower than other Jewish Israeli households, and over the last five years there was a 33% increase in the number of male yeshiva and kollel students in Israel further underlining the necessity of women to be breadwinners. In the last decade a rise was recorded in the percentage of ultra-Orthodox girls taking Bagrut (matriculation) exams (from 31% to 55%), compared to a decline in the parallel percentage among ultra-Orthodox boys (from 16% to 13%).

An analysis of the qualitative interviews elicited various antecedents for selecting to acquire an academic diploma (most participants noted more than one), which may be classified under 3 categories: breaking the community societal boundaries and building connections with outer-world, unlocking potential for self-fulfilment as an ultra-orthodox woman, and materialistic sustenance to escape the poverty circle as sole-provider since husband is unemployed.

These long-term life-changing benefits sourcing from freedom of career related choice were coupled with risks including: decline in faith, breaking ties with family and friends, and fear of getting stuck in the middle, neither being accepted by secular community to find a satisfying job opportunity nor able to fit anymore into the religious community identity after being exposed to non-religious contents and temptations. Above all, the goal-oriented ultra-orthodox students were anxious that their community leadership can order their secretive expulsion from religious society circles (daven, arranged-marriage, etc.) by forbidding or limiting contact with them, out of fear their conventions cause to lose control of rabbis over members by diffusing secular norms of behavior within community.

The women expressed notable challenges such as study overload (cause by inadequate English proficiency and poor mathematical background) which was difficult to balance with family and kids. For example, having classes on Friday morning can be of particular challenge for an ultra-orthodox woman who needs to cook meals for Shabbat (or high holidays) and has to halt all activities in the afternoon before shabbat enters. Interestingly, the quarantine period of COVID-19 in which the college transferred to online paradigm proved to be convenient for the ultra-orthodox track and it is going to be continued after the pandemic is over.

After analysing the data, we postulate that students' concerns are far-reaching than academic skills extending to mental capacities, since the orthodox system of education undermines critical thinking and disobedience to authority which are required in higher education to pursue innovation and entrepreneurship skills. For example, the libraires in ultra-orthodox community blocks internet connection and forbids offering liberal content by that any secular public library archive holds. Having an all-women orthodox college program helped to mitigate concerns by decreasing fears of alienation, prejudice or discrimination. Nevertheless, some ultraorthodox students felt pioneers who can bring progress to their home-front by becoming breadwinners. Their determination to succeed in the expense of personal cost is admirable.

From the data collection emerged a silhouette concurring with the literature review of an ultraorthodox student hero's journey (Campbell, 1949), traveling outside it her enclave community comfort zone, undergoing an unlearning process followed by a brand-new knowledge acquisition which is subsequently shared with her family and friends opening-up the community to the 21st century technological progress and cultural mosaic.

Our data analysis highlighted several facts: Orthodox women hear about opportunity to study from friends in the community. Secular women receive information from media advertisement. Orthodox women have weak training in mathematics, physics and English. In the area of computer sciences there was no differences between orthodox and secular women. The drop-out rate in Azrieli general college is about 30% whereas in orthodox Tmura college the drop-out rate is 1/40, indeed significantly lower serving as evidence for the success of the Ultra-orthodox seminary tailored for the students' needs based on their lower preparation and community lack of resources. Orthodox colleges foster success by creating supportive religious environment and extending scholarships because most orthodox students are unemployed in contrast to secular students who tend to work during learning period.

Implications for Practice

In the current research endeavour, we have integrated qualitatively derived themes with quantitative findings to depict the experiences of ultra-Orthodox female students in Machon Tal an Tmura as an exemplar of ultra-orthodox women colleges teaching STEM towards Bachelor and Master degrees in the city of Jerusalem, Israel. Table 3, summarizes the information accumulated from the interviews and secondary data sources about Machon Tal and Tmura.

Table 3: Comparison of machon Tal and Tmura

	JCE- Machon Tmura	**JCT – Machon Tal**
Year of establishment	2011	1990
Degrees offered	Industrial engineering, software engineering	Industrial Engineering, Computer Sciences, Bio-Informatics, Electro-optics, Business Management, Nursing
Type of students	Orthodox women only	Orthodox and conservative women (Dati Leumi)
Mechina preparation	6 months	From 3 to 12 months
Entry requirement	Matriculation certificate (Bagrut), Psychometry	Matriculation certificate (Bagrut), Psychometry, Interview
Years to graduate	4	3-5
Diversity of classes	Bible studies	Bible studies
Placement of alumni	Intel, Mobileye, Cinamedia, government agencies	El Al airline, Ophir optronics, NDS, Civan, Hadassah hospital
Location	Jerusalem	Jerusalem
Dormitory	No	Yes
Day, evening, classes	Day and evening	Day
Online classes	No	Yes
English Program	No	Yes
Number of students	40 per year	2000 overall
Director Name	Dini Weiss	Chaim Sukenik
Scholarships for student	Yes	Yes
Grant Funding sources	Keren Kemach, Toronto	Various
Tuition fee	Israel government standard academic university fee	Israel government standard academic university fee
Malag supervision	Yes	Yes
Level of Diploma	Undergraduate	Undergraduate and Master
Track for Ethiopian students	No	Yes

As a managerial outcome stemming from this study, we curate several useful practices that Machon Tmura (JCE) can potentially emulate from Machol Tal (JCT) which have proved fruitful to smooth the academic journey of Ultra-orthodox women:

1) Start an English track for ultra-orthodox women since in Jerusalem there are many new immigrants from English speaking countries.
2) The ultra-orthodox women need extensive scholarships since their economic status is poor. The college should maybe find the students part time jobs such as mentoring kids in their neighborhood (Perach program - flower in Hebrew language).
3) There needs to be mentorship program for students especially in the first year to decrease amount of dropout. The mentors should be advanced year students or alumni.
4) JCE needs to emulate the successful JCT program for preparation of students to academic studies called in Hebrew, Mechina, because many ultra-orthodox women arrive from high schools with weak matriculation tracks (called Bagrut in Hebrew) in Mathematics, Chemistry, Physics, etc.
5) A Nursing program in JCE similar to the one conducted at JCT can become a great success in JCE too since Jerusalem area has several large hospitals (Shari Tzedek and Hadassah) that suffer from shortage of nurses.
6) JCE needs to encourage women students from their second or third year of study to start having part time internships in Hi-Tech companies in Jerusalem area. Internships are great way to network for job search.
7) JCE should expand offering of online courses that will allow orthodox women to balance family life with education since they have many kids at home that need attention while spouses are in Bible study sessions.
8) JCE should put emphasis on increasing diversity by enrolling women from Ethiopian roots. This will contribute to create atmosphere of inclusion on campus.
9) JCE should provide dormitories for orthodox women that fit their level of observance. This will allow students to focus on their studies with distraction since families live in small apartments with ten or more siblings' members.

CONCLUSION

This study is one of the first of its kind to analyse a unique group of minority students, ultra-orthodox women from the Haredi community in Jerusalem, Israel, who are embarking on a journey to gain an academic diploma specifically in STEM, whereas most past studies focused on social work disciplines. While past studies of this sort tend to focus on college internal practical religious accommodations such as holidays vacations, prayer options, dietary needs and religious clubs (Ahmad, 2001; Ali and Bagheri, 2009; Bowman and Smedley, 2013; Cole and Ahmadi, 2003; 2010; Modood, 2006; Nasir and Al-Amin, 2006), our study took a different approach by fine-graining the student struggle with her community as consequence of breaking traditions. We suggest acknowledging differences between communities rather than minimizing them within college by trying to build a one size fit all curriculum which doesn't recognize the identity conflicts of highly religious students in college. There is a need to diminish fear among prospective students in the admission orientations that the college isn't going to contest their spiritual beliefs or erode their faith in God. There is a body of literature about cultural shock of first-generation students which is relevant to Ultra-orthodox segment too (McCoy, 2014; Moschetti and Hudley, 2015). The ultra-orthodox student journey from enclave community to college can be viewed as parallel to an international student who travels to study abroad in a foreign land with different dominant religion (Zhou et al., 2008).

The data elicited an admirable character of student who bravely steps-out of their community comfort-zone to become a breadwinner for their household while putting at high stake their social status within the community. An insight emerging from the study is the ability of ultra-orthodox students to graduate successfully with low drop-out rate. Importantly, the students were able to find a rewarding job proving their personal sacrifice was justified and setting a positive example for followers. This accomplishment should not be taken for granted as the Hi-tech market in Israel is competitive with clear preference to hire alumni from the Army prestige technological units, an experience that the ultra-orthodox students are unable to highlight in their curriculum vitae as many were lawfully exempted from service due to their religious/marital status. This successful outcome should be attributed the Azrieli College staff who tailored a program with curated curriculum which meets the students' needs not only academically but also culturally. The study has high value for policymakers in the government education ministry since the ultra-orthodox community size is growing has been exponentially in Israel because of high birth-rate adherence to the community traditions. We believe that the ultra-orthodox students should be viewed as agents of change who have the potential to bring back to their community new paradigms of thought which may ultimately break silos between parties in the Israeli society, serving a critical role in reshaping the state of Israel mosaic and meeting political needs for social dialogue to bridge gaps as evidenced in five counter-productive elections within three years (2019-2021).

Two fruitful avenues exist for future research. First, assessing whether results achieved in this study apply to ultra-orthodox Jewish women outside Israel borders. For example, Longman (2008) started an attempt to addresses this question through an investigation conducted inside the Orthodox Jewish community of Antwerp, Belgium. This fascinating topic should be examined in a subtle way because the Jewish communities around the world are composed of multiple streams. For example, in United States of America there are reform, modern-orthodox and ultra-orthodox communities which deserve equal research attention. Finally, an opportunity to continue this stream of research would be to investigate whether the conclusions reached in this study apply internationally for colleges of other religions (i.e., Catholic, Muslim. etc.) since women encounter similar challenges concerning tension between family status and career across the aisle.

REFERENCES

Ahmad, F. 2001. Modern traditions? British Muslim Women and Academic Achievement. *Gender and Education* 13 (2): 137 – 152.

Ali, S. R., and E. Bagheri. 2009. " Practical Suggestions to Accommodate the Needs of Muslim Students on Campus." *New Directions for Student Services* 2009 (125): 47 – 54.

Almond, G. A., R. S. Appleby, and E. Sivan. 2003. *Strong Religion: The Rise of Fundamentalisms Around the World.* Chicago: University of Chicago Press.

Ammerman, N. T. 2013. *Sacred Stories, Spiritual Tribes: Finding Religion in Everyday Life.* Oxford: Oxford University Press.

Armstrong, M. A., & Jovanovic, J. (2015). Starting at the crossroads: Intersectional approaches to institutionally supporting underrepresented minority women STEM faculty. *Journal of Women and Minorities in Science and Engineering*, *21*(2).

Astin, A. W., H. S. Astin, and J. Lindholm. 2011. *Cultivating the Spirit: How College can Enhance Students' Inner Lives*. San Francisco, CA: Jossey-Bass.

Aud, S., M. A. Fox, and A. Kewal Ramani. 2010. *Status and Trends in the Education of Racial and Ethnic Groups.* NCES 2010–2015. Washington, DC: National Center for Education Statistics.

Baum, N., T. Yedidya, C. Schwartz, and O. Aran. 2014. " Women Pursuing Higher Education in Ultra-Orthodox Society." *Journal of Social Work Education* 50 (1): 164 – 75.

Bowman, N. A., and J. W. Brandenberger. 2012. " Experiencing the Unexpected: Toward a Model of College Diversity Experiences and Attitude Change." *The Review of Higher Education* 35 (2): 179 – 205.

Bowman, N. A., and C. Smedley. 2013. " The Forgotten Minority: Examining Religious Affiliation and University Satisfaction." *Higher Education: The International Journal of Higher Education and Educational Planning* 65 (6): 745 – 60.

Bryant, A. N. 2006. " Exploring Religious Pluralism in Higher Education: Non-Majority Religious Perspectives among Entering First-Year College Students." *Religion & Education* 33 (1): 1 – 25.

Bryant, A. N. 2007. " Gender Differences in Spiritual Development During the College Years." *Sex Roles* 56 (11–12): 835 – 46.

Cahaner, L. (2011). *Proposal to establish the Matanel Career Center for the Ultra-orthodox (Haredi) Community at the University of Haifa*. Haifa: Haifa university.

Campbell, J. 1949. *The Hero with a Thousand Faces*. Princeton: Princeton University Press.

Carpenter, J. A. 1998. " Religion in American Academic Life." *Religion and American Culture: A Journal of Interpretation* 8 (2): 265 – 81.

CBS (Central Bureau of Statistics Israel). 2011. "Society and Population." CBS. http://www.cbs.gov.il/reader/?MIval = cw%5fusr%5fview%5fSHTML&ID=705.

CBS (Central Bureau of Statistics Israel). 2017 June. "Press Release 185/2017." CBS. Accessed January 11 2017. http://www.cbs.gov.il/reader/pdf.185e%5f17%5f16/hodaot2017n.

Cole, D., and S. Ahmadi. 2003. " Perspectives and Experiences of Muslim Women who Veil on College Campuses." *Journal of College Student Development* 44 (1): 47 – 66.

Cole, D., and S. Ahmadi. 2010. " Reconsidering Campus Diversity: An Examination of Muslim Students' Experiences." *The Journal of Higher Education* 81 (2): 121 – 39.

Dehan, N. (2013). Case study: The first Israeli BSW program for Haredi women students. *Journal of Social Work Education*, *49*(4), 716-732.

Dey, E. L., M. C. Ott, M. Antonaros, C. Barnhardt, and M. A. Holsapple. 2010. *Engaging Diverse Viewpoints: What is the Campus Climate for Perspective-Taking*? Washington, DC: Association of American Colleges and Universities.

Eisenstadt, S. N. 2000. " Multiple Modernities." *Daedalus* 129 (1): 1 – 29.
Fischer, M. J. 2007. " Settling Into Campus Life: Differences by Race/Ethnicity in College Involvement and Outcomes." *Journal of Higher Education* 78 (2): 125 – 61.
Fowler, J. W. 2004. " Faith Development at 30: Naming the Challenges of Faith in a New Millennium." *Religious Education* 99 (4): 405 – 21.
Garr, M., & Marans, G. (2001). Ultra-orthodox women in Israel: A pilot project in social work education. *Social work education*, *20*(4), 459-468.
Geiger, B., & Alt, D. (2014). Haredi/Chabad Women's Acculturation Experience in a Non-Haredi Institute of Higher Education in Northern Galilee. *Women in Judaism: A Multidisciplinary e-Journal*, *11*(1).
Getz, D., Shacham, O. K., & Klein, R. (2019). *Women in Technological and Scientific Entrepreneurship.*
Golan, O., & Fehl, E. (2020). Legitimizing academic knowledge in religious bounded communities: Jewish ultra-orthodox students in Israeli higher education. *International Journal of Educational Research*, *102*, 101609.
Golan, O., & Campbell, H. A. (2015). Strategic management of religious websites: The case of Israel's orthodox communities. *Journal of Computer-Mediated Communication*, *20*(4), 467-486.
Gottlieb,D.(2007). *Poverty and labor market behavior in the UltraOrthodox com munity.* Jerusalem: Van Leer Institute.
Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters?. *Economics of Education Review*, *29*(6), 911-922.
Gurovich, N., & Cohen-Kastro, E. (2004). *Ultra-Orthodox Jews geographic distribution and demographic, social and economic characteristics of the Ultra-Orthodox Jewish population in Israel 1996-2001.* Jerusalem: Central Bureau of Statistics – Demography Sector.
Hakak, Y. (2004). *Between the holy and the purposeful: Vocational training for Ultra-Orthodox men*. Jerusalem: The Floersheimer Institute for Policy Studies.
Hartley, H. V. 2004. " How College Affects Students' Religious Faith and Practice: A Review of Research." *College Student Affairs Journal* 23 (2): 111 – 29.
HERI (Higher Education Research Institution). 2005. *The Spiritual Life of College Students: A National Study of College Students' Search for Meaning and Purpose*. Los Angeles: University of California, Higher Education Research Institution.
Herzog, H., and A. Braude, eds. 2009. *Gendering Religion and Politics: Untangling Modernities.* New York: Palgrave Macmillan.
Hleihel, A. 2011. *Fertility among Jewish and Muslim Women in Israel, by Level of Religiosity,* 1979–2009. Jerusalem: State of Israel Central Bureau of Statistics (Hebrew).
Hurtado, S. 2007. " Linking Diversity with the Educational and Civic Missions of Higher Education." *The Review of Higher Education* 30 (2): 185 – 96.
Hyers, L. L., and C. Hyers. 2008. " Everyday Discrimination Experienced by Conservative Christians at the Secular University." *Analyses of Social Issues and Public Policy* 8: 113 – 37.
Israel Advanced Technology Industries (2020). The Haredi High-tech report 2020, Available at: https://www.iati.co.il/files/files/THE%20HAREDI%20HI-TECH%20REPORT%202020.pdf
Israel Democracy Institute (2020). Employment rates for ultra-orthodox women continue to rise, Published on December 29, available at: https://en.idi.org.il/articles/33344
Jacobsen, D., and R. H. Jacobsen, eds. 2008. *The American University in a Postsecular Age: Religion and Higher Education*. New York: Oxford University Press.
Kalaagi, T. 2012. " *Between the Wig and the Hijab: A Comparative Look at Academization among Women from Conservative Societies – Haredi and Bedouin in Israel.*" Doctoral diss., Ramat Gan: Bar Ilan University. (Hebrew).

Laiush, B. 2014. *The Liminal Women: Haredi Women Facing Modern Change.* Tel Aviv : Resling. (Hebrew).
Lee, J. J. 2002. " Religion and College Attendance: Change among Students." *Review of Higher Education* 25: 369 – 84.
Longman, C. (2008). Sacrificing the career or the family? Orthodox Jewish women between secular work and the sacred home. *European Journal of Women's Studies*, *15*(3), 223-239.
Odenheimer, A., & Ackerman. G. (2012, June 28). Israel existential crisis creates Haredi breadwomen as men study. Retrieved from http://www.bloomberg.com
Malach, G. 2014. A "Kosher" Degree – Academic Studies in the Haredi Sector. Jerusalem: Floersheimer Studies, The Hebrew University of Jerusalem (Hebrew).
Malach, G., L. Kahner, and E. Regev. 2016. *The Five-Year Plan of MALAG-VATAT for the Haredi Population Over 2012–2016, Assessment and Recommendation for the Next Five Years.* Jerusalem: MALAG-VATAT. (Hebrew).
Malhi, A., Cohen, B., & Kaufmann, D. (2008). *The Ultra-Orthodox: Fearing for their future – Positions and obstacles relating to higher education in the UltraOrthodox Sector.* Jerusalem: The Jerusalem Institute for Israel Studies.
Mayhew, M. J., N. A. Bowman, and A. B. Rockenbach. 2014. " Silencing Whom? Linking Campus Climates for Religious, Spiritual, and Worldview Diversity to Student Worldviews." *The Journal of Higher Education* 85 (2): 219 – 45.
Mayrl, D., and F. Oeur. 2009. " Religion and Higher Education: Current Knowledge and Directions for Future Research." *Journal for the Scientific Study of Religion* 48 (2): 260 – 75.
McCoy, D. 2014. " A Phenomenological Approach to Understanding First-Generation College Students of Color Transitions to One "Extreme" Predominantly White Institution." *College Student Affairs Journal* 32 (1): 155 – 69.
Meeuwisse, M., S. E. Severiens, and M. P. Born. 2010. " Reasons for Withdrawal from Higher Vocational Education. A Comparison of Ethnic Minority and Majority non-Completers." *Studies in Higher Education* 35 (1): 93 – 111.
Modood, T. 2006. " Ethnicity, Muslims and Higher Education Entry in Britain." Teaching in Higher Education 11 (2): 247 – 50.
Mooney, M. 2010. " Religion, College Grades, and Satisfaction among Students at Elite Colleges and Universities." *Sociology of Religion* 71 (2): 197 – 215.
Moschetti, R. V., and C. Hudley. 2015. " Social Capital and Academic Motivation among First-Generation Community College Students." *Community College Journal of Research & Practice* 39 (3): 235 – 51.
Mubarak, H. 2007. "*How Muslim Students Negotiate their Religious Identity and Practices in an Undergraduate Setting [Electronic Version]*." Essay Forum on the Religious Engagements of American Undergraduates, 1–12. Accessed July 18, 2017 http://religion.ssrc.org/reforum/.
Nasir, N. I. S., and J. Al-Amin. 2006. " Creating Identity-Safe Spaces on College Campuses for Muslim Students." *Change: The Magazine of Higher Learning*, 38 (2), 22 – 27.
Novis Deutsch, N., and C. Lifshitz. 2016. " Integrating Conservative Religious Minorities into Higher Education: Praxis, Identity and Value Challenges of Teaching Ultra-Orthodox Female Religious Students from the Teacher's Perspective." *Teaching in Higher Education* 16 (9): 485–500.
Novis Deutsch, N., & Rubin, O. (2019). Ultra-Orthodox women pursuing higher education: motivations and challenges. *Studies in Higher Education*, *44*(9), 1519-1538.
Orbe, M. P. 2008. " *Theorizing Multidimensional Identity Negotiation: Reflections on the Lived Experiences of First-Generation College Students.*" New Directions for Child and Adolescent Development 120: 81 – 95.
Parks, S. D. 2000. *Big Questions, Worthy Dreams: Mentoring Young Adults in Their Search for Meaning, Purpose and Faith.* San Francisco: Jossey-Bass.

Pascarella, E. T., and P. T. Terenzini. 2011. *How College Affects Students: A Third Decade of Research,* The Jossey-Bass higher and adult education series.
Pfefferman, B., and A. Malchi. 2010. *Employment in the Haredi Sector – Special Report. Jerusalem: The Israeli ministry of Industry, Trade and Labor.* (Hebrew).
Regev, E. 2016. "*Challenges in the Integration of Haredi Jews into Academic Studies*." Report on the state of the nation: Society, economy and policy. The Taub center for research in social policy in Israel (Hebrew).
Reimer, S. 2010. " Higher Education and Theological Liberalism: Revisiting the old Issue." *Sociology of Religion* 71 (4): 393 – 408.
Rubin, O., and N. Novis-Deutsch. 2017. " Profiles of Haredi Women Pursuing Higher Education: A Cluster Analysis of Diverse Patterns of Motivations and Concerns." Megamot: *Journal for Behavioral Sciences* 52 (2): 109–140. (Hebrew).
Schmalzbauer, J. 2013. " Campus Religious Life in America: Revitalization and Renewal." *Society* 50 (2): 115 – 31.
Schmalzbauer, J., and K. Mahoney. 2012. "*Religion and Knowledge in the Post-Secular Academy.*" In The Post-Secular in Question: Religion in Contemporary Society, edited by P. Gorski, D. K. Kim, J. Torpey, and J. VanAntwerpen, 215 – 248. New York, NY: New York University Press.
Sherkat, D. E. 2007. "*Religion and Higher Education: The Good, the Bad, and the Ugly*." SSRC. Accessed July 9 2016. http://religion.ssrc.org/reforum/Sherkat.pdf.
Siddiqi, I. April 7, 2016. "*Here are 6 Incidents of Islamophobia on Campus in the Last Week.*" Mondoweiss. Accessed October 30 2017. http://mondoweiss.net/2016/04.
Sivan, E., and K. Caplan. 2003. *Israeli Haredim: Integration Without Assimilation*. Tel Aviv: Hakibbutz Hameuhad (Hebrew).
Small, J., and N. Bowman. 2011. " Religious Commitment, Scepticism, and Struggle among US College Students: The Impact of Majority/Minority Religious Affiliation and Institutional Type." *Journal for the Scientific Study of Religion* 50: 154 – 74.
Thomas, G. D. 2001. " *The Dual Role of Scholar and Social Change Agent: Reflections from Tenured African American and Latina Faculty."* In Sisters of the Academy: Emergent Black Women Scholars in Higher Education, edited by R. O. Mabokela, and A. L. Green, 80 – 91. Sterling, VA: Stylus.
Tinto, V. 2012. *Completing College: Rethinking Institutional Action.* Chicago, IL: University of Chicago Press.
Tobin, G. A., A. K. Weinberg, and J. Ferer. 2009. *The Uncivil University: Intolerance on College Campuses.* Lanham, MD: Lexington Books.
Yin, R.K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, California: Sage Publications.
Zhou, Y., D. Jindal-Snape, K. Topping, and J. Todman. 2008. " Theoretical Models of Culture Shock and Adaptation in International Students in Higher Education." *Studies in Higher Education* 33 (1): 63 – 75.

DECISION SCIENCES INSTITUTE

Multi-Period Set Covering Problems, Application, and Solution

Hadi Farhangi
Savannah State University
Email: farhangih@savannahstate.edu

ABSTRACT

Set covering problem is a combinatorial optimization problem that selects sets to cover all elements of a universal set. Set covering problem has many applications in scheduling and systems design. In this paper, we extend the set covering problem to multiple periods of planning. The extension is not unique, in which a decision maker may face different needs which may lead to various types of extensions. For example, one may select no more than one set in each period or one may solve a single set covering in each period. Nevertheless, various of extensions and their formulations are presented in this paper. We also discuss one application for multi-period set covering problems. One of the variations of the problem is solved using Cplex to provide insight on the computational complexity of the problem.

KEYWORDS: Set Covering, Multi-Period, System of Systems

INTRODUCTION

Set Covering (SC) problem is one of the basic combinatorial problems that cannot be reduced into an easier problem. This problem has a variety of applications in the industry such as military (Konur et al., 2016; Farhangi and Konur, 2018; Argun, 2019) and crew scheduling (Caprara et al., 1999) among many others. The presented work focuses on extending the set covering problem into multiple periods of planning. The extensions provide opportunity for the planning of multiple periods problems that are based on SC problems. Particularly, System of Systems (SoS) architecting can benefit from the extensions as it does not, currently, account for the temporary and evolutionary properties of SoS (Farhangi and Konur, 2018).

To state the set covering problem, we use a similar notation that appears in (Lan et al., 2007; Farhangi et al., 2016). Let the set of items of the universal set be indexed by $j \in J = \{1, ..., |J|\}$ and the set of sets to cover the universal set be indexed by $i \in I = \{1, ..., |I|\}$. Note that we have inequality $|I| \leqslant 2^{|J|} - 1$ between sets I and J. Set covering problem has a parameter $a_{ji} = 1$, which means set $i \in I$ has the element $j \in J$ and $a_{ji} = 0$, otherwise. In addition, c_i is the cost of adding a set. We also define the variable $x_i = 1$ if system i is selected and $x_i = 0$, otherwise. Then, the set covering problem is:

$$\begin{aligned}
\text{SC:} \quad \min \quad & \sum_{i \in I} c_i x_i & & (1)\\
\text{s.t.} \quad & \sum_{i \in I} a_{ji} x_i \geqslant 1 & \forall j \in J \quad & (2)\\
& x_i \in \{0,1\} & \forall i \in I \quad & (3)
\end{aligned}$$

In this formulation, equation (1) is the total cost of selecting sets. Constraints (2) are covering constraints which guarantee the covering of the elements of the universal set. Constraints (3) impose the binary definition for the variables.

In SC problem, we plan for one period and decide on the optimum selection of sets satisfying the covering requirement of the problem. For the extensions, we study multiple periods of planning, in which, we plan for multiple periods of a finite planning horizon and select sets satisfying the requirement of the problem in each period. These extensions are important as they can capture the dynamic of the environment following the work of (Farhangi and Konur, 2018).

Surveys related to multiple periods planning and covering requirements of SC problem can be seen in many maximal covering works such as Li et al. (2011). A related work to the MPSC problem is the work of Chrissis et al. (1982). Particularly, Chrissis et al. (1982) restate the covering constraints for every period of planning, which we did not study and we focus on other variations for our extensions. The first contribution of this paper lies in the formulation of the various extensions of the SC problem. We present constraints similar to the set covering constraints in Farhangi and Konur (2018), maximal covering constraints in Rajagopalan et al. (2008); Li et al. (2011); Sha and Huang (2012); Benneyan et al. (2012); Belanger et al. (2015); Colombo et al. (2016), and multiple periods constraints in Chrissis et al. (1982). To the best of our knowledge, four of our presented formulations are new in the literature.

MPSC problem has a variety of applications, among which one can name the temporary and evolution properties of the SoS. As noted in Farhangi and Konur (2018), these characteristics of the SoS is not formulated mathematically, which is the second contribution of this paper. The organization of the paper is as follows: the extensions of SC problems are presented, next. We discuss applications thereafter along with an analysis of solutions. The paper is concluded in the last section.

MULTI-PERIOD SET COVERING PROBLEMS

In this section, we present a variety of formulations to extend the SC problem. We have a set of periods that is indexed by $t \in T = \{1, ..., |T|\}$ and we extend SC problem to every period $t \in T$, and we call it Multi-Period Set Covering (MPSC) problem. The extension affects sets I, J and parameters a_{ji} and c_i. We define I^t and J^t for each period. In addition, we redefine a^t_{ji} and c^t_i for each period and they can change in some or all periods. We make the extensions based on our knowledge of the sets I^t, the change in parameters a^t_{ji}, c^t_i, and the limitation of the available budget. Note that for these extensions, we define variable $x^t_i = 1$ if set $i \in I$ is selected in period $t \in T$ and $x^t_i = 0$, otherwise.

MPSC problems with Known I^t and J^t

In this extension, sets $I^t \forall t \in T$ are known and there may or may not be a relation between sets I^t and I on one hand and J^t and J on the other hand. This extension can be expressed as problem *MPSC1*.

$$\textit{MPSC1}: \quad \min \sum_{t \in T} \sum_{i \in I^t} c^t_i x^t_i \tag{4}$$

$$\text{s.t.} \quad \sum_{i \in I^t} a_{ji}^t x_i^t \geqslant 1 \quad \forall j \in J^t, \forall t \in T \quad (5)$$
$$x_i^t \in \{0,1\} \quad \forall i \in I^t, \forall t \in T \quad (6)$$

In this formulation, equation (4) is the total cost of selecting sets in all periods. Constraints (5) are covering constraints which translate to the covering elements of the universal sets J^t in each period. Constraints (6) impose the binary definition for the variables.

MPSC problems with Unknown I^t

For the case of unknown J^t the formulation is undecided as the universal sets to be covered are unknown. For this reason, we consider the case that the sets I^t are unknown. We refer to this extension as *MPSC2*. This case arises particularly when we have capacity constraints or budget constraints and we cannot select all the required sets in each period. To formulate *MPSC2* problem, we define variable $y_j^t = 1$ if item $j \in J$ is selected in period $t \in T$, otherwise $y_j^t = 0$. Ideally, one needs to maximize the number of covered items of the universal set, J, in each period. Assuming that the f_j^t is the cost of having element j in period t, h_i^t is the cost of including set i in period t, and g^t is the total available budget in period $t \in T$, we can formulate *MPSC2* as follows:

$$\textit{MPSC2}: \quad \max \quad \sum_{t \in T} \sum_{j \in J} y_j^t \quad (7)$$
$$\text{s.t.} \quad \sum_{i \in I} a_{ji}^t x_i^t \geqslant y_j^t \quad \forall j \in J, \forall t \in T \quad (8)$$
$$\sum_{t \in T} \sum_{i \in I} a_{ji}^t x_i^t \geqslant 1 \quad \forall j \in J \quad (9)$$
$$\sum_{i \in I} h_i^t x_i^t + \sum_{j \in J} f_j^t y_j^t \leqslant g^t \quad \forall t \in T \quad (10)$$
$$x_i^t \in \{0,1\} \quad \forall i \in I, \forall t \in T \quad (11)$$
$$y_j^t \in \{0,1\} \quad \forall j \in J, \forall t \in T \quad (12)$$

Equation (7) enforces the model to maximize the summation of selected items in all periods. Constraints (8) are covering constraints for the selected items in each period. Particularly, if $y_j^t = 1$ the covering constraints (8) will be active. Constraints (9) guarantee that all elements of the universal set will be covered in the planning horizon. Constraints (10) enforce an upper bound on the spent budget in each period. Finally, constraints (11) and (12) are the definition of variables.

In problem *MPSC2*, we maximize the selection of items $j \in J$ over all periods. Alternatively, one may need to maximize the minimum number of selected items in each period. In another word, we can consider the following max-min expression: $\max_{t \in T}\{\sum_{j \in J} y_j^t, \forall t \in T\}$. The resulting model is called *MPSC3* and it is stated in the following as a mixed-integer linear programming model:

$$\textit{MPSC3}: \quad \max \quad \lambda \quad (13)$$
$$\text{s.t.} \quad \sum_{j \in J} y_j^t \geqslant \lambda \quad \forall t \in T \quad (14)$$

$$\sum_{i\in I} a_{ji}^t x_i^t \geqslant y_j^t \qquad \forall j \in J, \forall t \in T \qquad (15)$$
$$\sum_{t\in T}\sum_{i\in I} a_{ji}^t x_i^t \geqslant 1 \qquad \forall j \in J \qquad (16)$$
$$\sum_{i\in I} h_i^t x_i^t + \sum_{j\in J} f_j^t y_j^t \leqslant g^t \qquad \forall t \in T \qquad (17)$$
$$x_i^t \in \{0,1\} \qquad \forall i \in I, \forall t \in T \qquad (18)$$
$$y_j^t \in \{0,1\} \qquad \forall j \in J, \forall t \in T \qquad (19)$$
$$\lambda \geqslant 0 \qquad (20)$$

Equation (13) and constraints (14) enforce the model to maximize the minimum number of selected items in each period, which are the linearization of the expression $\max\Big\{\min_{t\in T}\{\sum_{j\in J} y_j^t\}\Big\}$. Constraints (8) are covering constraints for the selected items in each period. Constraints (16) guarantee that all elements of the universal set will be covered in at least one period. Constraints (17) are budget constraints. Finally, constraints (18)–(20) are the definition of variables.

Problems *MPSC2* and *MPSC3* enforce budget constraints for each period. Instead, one can consider a budget constraint for the entire planning horizon with the available budget equal to G. This requirement is shown in the following problem, which is a binary model.

MPSC4:
$$\max \quad \sum_{t\in T}\sum_{j\in J} y_j^t \qquad (21)$$
$$\text{s.t.} \quad \sum_{i\in I} a_{ji}^t x_i^t \geqslant y_j^t \qquad \forall j \in J, \forall t \in T \qquad (22)$$
$$\sum_{j\in J} y_j^t \geqslant 1 \qquad \forall t \in T \qquad (23)$$
$$\sum_{t\in T}\sum_{i\in I} h_i^t x_i^t + \sum_{t\in T}\sum_{j\in J} f_j^t y_j^t \leqslant G \qquad (24)$$
$$x_i^t \in \{0,1\} \qquad \forall i \in I, \forall t \in T \qquad (25)$$
$$y_j^t \in \{0,1\} \qquad \forall j \in J, \forall t \in T \qquad (26)$$

Equation (21) enforces the model to maximize the summation of selected items in all periods. Constraints (22) are covering constraints for the selected items in each period. Particularly, if $y_j^t = 1$ the covering constraints (22) will be active. Constraints (23) guarantee that at least one element will be covered in each period. Constraints (24) enforce an upper bound on the spent budget in the planning horizon. Finally, constraints (25) and (26) are the definition of variables.

It is worth noting that one may decide to maximize the minimum number of selected elements in each period in the setting of problem *MPSC4*. In that case, the new problem will take the following form which we call *MPSC5*.

MPSC5:
$$\max \quad \lambda \qquad (27)$$
$$\text{s.t.} \quad \sum_{j\in J} y_j^t \geqslant \lambda \qquad \forall t \in T \qquad (28)$$

$$\sum_{i \in I} a_{ji}^t x_i^t \geqslant y_j^t \qquad \forall j \in J, \forall t \in T \qquad (29)$$

$$\sum_{j \in J} y_j^t \geqslant 1 \qquad \forall t \in T \qquad (30)$$

$$\sum_{t \in T} \sum_{i \in I} h_i^t x_i^t + \sum_{t \in T} \sum_{j \in J} f_j^t y_j^t \leqslant G \qquad (31)$$

$$x_i^t \in \{0,1\} \qquad \forall i \in I, \forall t \in T \qquad (32)$$

$$y_j^t \in \{0,1\} \qquad \forall j \in J, \forall t \in T \qquad (33)$$

$$\lambda \geqslant 0 \qquad (34)$$

Equation (27) and constraints (28) enforce the model to maximize the minimum number of selected items in each period, similar to the problem *MPSC3*. Constraints (29) are covering constraints for the selected items in each period. Constraints (30) guarantee that all elements of the universal set will be covered in at least one period. Constraints (31) are budget constraints for the entire planning horizon. Finally, constraints (32)–(34) are the definition of variables.

APPLICATION

This paper develops *MPSC* problems for application purposes such as architecting SoS (system of systems) for multiple periods of planning. SoS concepts were developed during the last 4 decades (Jackson and Keys, 1984) to address arising complex issues (Agarwal et al., 2014; Jamshidi, 2017). Its usage spans the military, space exploration, powergrid, healthcare, national security, and many others (Sage, 2011; DeLaurentis, 2005; Ender et al., 2010; Dahmann and Baldwin, 2008). The architecting of SoS is based on five properties (Farhangi and Konur, 2018); (i) it is capable, (ii) it creates connection between systems, (iii) it evolves over time, (iv) it is available temporary, and (v) it exhibits emergent behavior. Currently, properties (iii) and (iv) are not included in the formulation of SoS (Farhangi et al., 2016; Farhangi and Konur, 2018). This gap can be partially addressed using *MPSC* problems.

To formulate this SoS problem considering properties (i)-(iv), we use notation similar to *MPSC* problems to ease the development of SoS. We assume systems are indexed by $i \in I$, capabilities are indexed by $j \in J$, and finite periods of planning are indexed by $t \in T$. We define problem parameters as follows; h_i^t is the cost of including system i in SoS in period t, f_j^t is the cost of providing capability j for SoS in period t, $c_{i_1 i_2}^t$ is the cost of connecting system i_1 to system i_2 in period t, g^t is the available budget in period t, and $a_{ji}^t = 1$ if system i can provide capability j in period t, otherwise $a_{ji}^t = 0$. We define variables as follows: $x_i^t = 1$ if system i is included in period t, $y_j^t = 1$ if capability j is provided in period t, and $z_{i_1 i_2}^t = 1$ if system i_1 is connected to system i_2 in period t. Now, SoS architecting problem can be formulated as follows:

SoS: $$\max \sum_{t \in T} \sum_{j \in J} y_j^t \qquad (35)$$

$$\text{s.t.} \quad \sum_{i \in I} a_{ji}^t x_i^t \geqslant y_j^t \qquad \forall j \in J, \forall t \in T \qquad (36)$$

$$\sum_{t \in T} \sum_{i \in I} a_{ji}^t x_i^t \geqslant 1 \qquad \forall j \in J \qquad (37)$$

$$\sum_{i_1 \in I} \sum_{i_2 \in I} c^t_{i_1 i_2} z^t_{i_1 i_2} + \sum_{i \in I} h^t_i x^t_i + \sum_{j \in J} f^t_j y^t_j \leqslant g^t \quad \forall j \in J \tag{38}$$

$$z^t_{i_1 i_2} \geqslant x^t_{i_1} + x^t_{i_2} - 1 \quad \forall i_2 > i_1 \in I, \forall t \in T \tag{39}$$

$$x^t_i \in \{0,1\} \quad \forall i \in I, \forall t \in T \tag{40}$$

$$y^t_j \in \{0,1\} \quad \forall j \in J, \forall t \in T \tag{41}$$

$$z^t_{i_1 i_2} \in \{0,1\} \quad \forall i_2 > i_1 \in I, \forall t \in T \tag{42}$$

In problem *SoS*, the objective is to maximize total covered capabilities in (35). Constraints (36) are covering constraints for the selected capabilities in each period. Particularly, if $y^t_j = 1$ the covering constraints (36) will be active. Constraints (37) guarantee that all capabilities will be covered in the planning horizon. Constraints (38) impose an upper bound on the spent budget in each period. Constraints (39) guarantee there are connections between selected systems (Konur et al., 2016; Farhangi et al., 2016). Finally, constraints (40)–(42) define variables.

Problem *SoS* covers required capabilities due constraints (36)–(37); hence, property (i) is achieved through the above formulation. Property (ii) is satisfied by constraints (39). Property (iv) is also satisfied because planning periods are over a finite horizon T. This formulation, however, does not satisfy the evolution of SoS. One simple way to impose evolution is to add constraints such as the following to the problem:

$$y^t_j \geqslant y^{t-1}_j \quad \forall j \in J, \forall t \in T \backslash \{1\} \tag{43}$$

Constraints (43) guarantee that if an element is chosen in one period, it must be chosen in the subsequent periods as well. This means the evolution of SoS in respect to the selected capabilities in each period, in which the SoS capabilities cannot be worsened. Nevertheless, if one removes variable $z^t_{i_1 i_2} = 1$ and constraints (39) and (42) from the problem *SoS*, the problem will be reduced to *MPSC2*. One can architect a SoS such that all total available budgets cannot exceed a certain value. In that case, *SoS* problem will be reduced to *MPSC4*. Similar arguments can be made in favor of problems *MPSC3* and *MPSC5*. In another word, *MPSC* problems can be considered as the core to formulate SoS problem over a given planning horizon. Note that the solution methods for problem *SoS* is beyond the scope of this work and we leave them for future studies.

SOLUTION

This section analyzes the performance of Cplex solver. We perform this analysis only for *MPSC2* problem as a guide for computational studies. The rest of variations can be solved using Cplex in a similar manner as described in this section. However, this variation is the most simple one and if solver shows difficulty or high computational time, the rest of problems will likely to show a similar pattern. For this reason, this study focuses on *MPSC2* problem and the the findings of this section will be used to solve all these problems in the future studies.

All procedures and routines are implemented in Matlab 2016a and run on a computer with Intel Core i3 2.3 GHz and 6 GB RAM. Furthermore, IBM ILOG's 128 Cplex solver is used for solving instances directly. The notation cpu shows the computational time in seconds, $|var|$ shows the number of variables, and $|const|$ shows the number of constraints for the solved instances. In addition, z^* shows the optimal objective value; $z^* = NA$ if no optimal integer solution is found. Finally, we restrict the computational time of Cplex to 500 seconds. Cplex often stops at a time after this limit; hence, we report that computational time and we use $opt\%$ to report the percentage of instances when the optimum solution is reached.

For the problem parameters, we generate $h_i^t \in UI[5, 20]$, $f_j^t \in UI[2, 10]$, and $g^t = \frac{1}{3}\left(\sum_{i\in I} h_i^t + \sum_{j\in J} f_j^t\right)$, where $UI[a, b]$ represents a random integer between a and b. The array of a_{ji}^t is generated randomly as a binary array; for a given $\hat{t}$, every row of $a_{ji}^{\hat{t}}$ matrix is checked to have at least one value of 1 in that row to guarantee that the resulting model is feasible to cover all elements of set J.

We generate two classes of instances based on the size of instances, which are small size are medium size classes, using the above setting. The size of instances are related to the values of $|I|$, $|J|$, and $|T|$. In other words, We vary these values to get different instance sizes. For the class of small size instances, we vary the values of $|I|$, $|J|$, and $|T|$ between 20, 30, and 40. For each combination of $|I|$, $|J|$, and $|T|$ values, we solve five instances. The average computational times of Cplex (cpu), in seconds, is reported in Table 1 in the appendix. As it can be seen in this table, on average, it takes Cplex more than 1036 seconds to solve instances. In addition, in around 40% of instances, Cplex is unable to find an integer solution.

For the class of medium size instances, we vary the values of $|I|$, $|J|$, and $|T|$ between 60, 70, and 80. The rest of the setting is similar to the small size class. The average computational times of Cplex is reported in Table 2 in the appendix. This table shows that Cplex solves instances in more than 1,745 seconds. Moreover, Cplex is unable to find an integer solution within the recorded cpu time in more than 77% of instances.

As it can been seen in Tables 1 and 2, average computational time increases from over 1036 seconds to over 1745 seconds. Furthermore, for medium size instances (Table 2), in more than 66% of instances, Cplex is unable to find an integer solution within the recorded cpu time in more than 77% of instances. This shows the difficulty of problem *MPSC2* when the size of the problem increases. For this reason, it is recommended to use large scale optimization methods such as Lagrangian relaxation, column generation, and heuristics. These advance techniques will be employed to solve all *MPSC2*, *MPSC3*, *MPSC4*, and *MPSC5* instances in the future studies.

CONCLUSION AND FUTURE RESEARCH

This paper extends the SC problem to multiple periods of planning, called *MPSC* problems. The extensions are not unique and they depend on our knowledge of sets I, J, and T. Nonetheless, the formulation for five different extensions are presented in this paper. We also discuss an application of *MPSC* problems for formulating SoS architecting problems considering multiple periods of planning. We solved one variation of *MPSC* problems, namely *MPSC2*, using Cplex and study the computational efforts for solving instances of this problem. In the future, one can study the use of Lagrangian relaxation, column generation, and heuristics methods for solving variations of *MPSC* problems and multi-period SoS architecting problems.

REFERENCES

Agarwal, S., Pape, L. E., Kilicay-Ergin, N., and Dagli, C. H. (2014). Multi-agent based architecture for acknowledged system of systems. *Procedia Computer Science*, 28:1–10.

Argun, I. D. (2019). *An Overview on Set Covering Problems With a Focus on Military Applications*, pages 54–66. IGI Global.

Belanger, V., Ruiz, A., Soriano, P., and Lanzarone, E. (2015). The ambulance relocation and dispatching problem.

Benneyan, J. C., Musdal, H., Ceyhan, M. E., Shiner, B., and Watts, B. V. (2012). Specialty care single and multi-period location-allocation models within the veterans health administration. *Socio-economic planning sciences*, 46(2):136–148.

Caprara, A., Fischetti, M., and Toth, P. (1999). A heuristic method for the set covering problem. *Operations research*, 47(5):730–743.

Chrissis, J. W., Davis, R. P., and Miller, D. M. (1982). The dynamic set covering problem. *Applied Mathematical Modelling*, 6(1):2–6.

Colombo, F., Cordone, R., and Lulli, G. (2016). The multimode covering location problem. *Computers & Operations Research*, 67:25–33.

Dahmann, J. S. and Baldwin, K. J. (2008). Understanding the current state of us defense systems of systems and the implications for systems engineering. In *2008 2nd Annual IEEE Systems Conference*, pages 1–7. IEEE.

DeLaurentis, D. (2005). Understanding transportation as a system-of-systems design problem. In *43rd AIAA aerospace sciences meeting and exhibit*, page 123.

Ender, T., Leurck, R. F., Weaver, B., Miceli, P., Blair, W. D., West, P., and Mavris, D. (2010). Systems-of-systems analysis of ballistic missile defense architecture effectiveness through surrogate modeling and simulation. *IEEE Systems Journal*, 4(2):156–166.

Farhangi, H. and Konur, D. (2018). System of systems architecting problems: Definitions, formulations, and analysis. In *Procedia Computer Science*, volume 140, pages 29–36.

Farhangi, H., Konur, D., and Dagli, C. H. (2016). A separation method for solving multiobjective set covering problem. In Yang, H., Kong, Z., and Sarder, M., editors, *Proceedings of the 2016 IISE*.

Jackson, M. C. and Keys, P. (1984). Towards a system of systems methodologies. *Journal of the operational research society*, 35(6):473–486.

Jamshidi, M. (2017). *Systems of systems engineering: principles and applications*. CRC press.

Konur, D., Farhangi, H., and Dagli, C. H. (2016). A multi-objective military system of systems architecting problem with inflexible and flexible systems: formulation and solution methods. *OR spectrum*, 38(4):967–1006.

Lan, G., DePuy, G. W., and Whitehouse, G. E. (2007). An effective and simple heuristic for the set covering problem. *European journal of operational research*, 176(3):1387–1403.

Li, X., Zhao, Z., Zhu, X., and Wyatt, T. (2011). Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research*, 74(3):281–310.

Rajagopalan, H. K., Saydam, C., and Xiao, J. (2008). A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers & Operations Research*, 35(3):814–826.

Sage, A. P. (2011). *System of systems engineering: innovations for the 21st century*, volume 58. John Wiley & Sons.

Sha, Y. and Huang, J. (2012). The multi-period location-allocation problem of engineering emergency blood supply systems. In *Systems Engineering Procedia*, volume 5, pages 21–28.

APPENDIX: TABLES

Table 1: Performance of Cplex for the small size *MPSC2* instances

$\|T\|$	$\|J\|$	$\|I\|$	$\|var\|$	$\|const\|$	z^*	cpu	$opt\%$
20	20	20	800	440	NA	453.37	80
		30	1000	440	399.8	0.34	100
		40	1200	440	400	0.1	100
	30	20	1000	650	NA	936.14	60
		30	1200	650	563.6	520.46	100
		40	1400	650	599.8	0.73	100
	40	20	1200	860	NA	987.98	60
		30	1400	860	NA	2316.48	0
		40	1600	860	NA	1467.61	60
30	20	20	1200	650	530.8	17.69	10
		30	1500	650	600	0.49	100
		40	1800	650	600	0.18	10
	30	20	1500	960	NA	2343.42	0
		30	1800	960	NA	2032.9	40
		40	2100	960	898.8	1.22	100
	40	20	1800	1270	NA	1918.72	20
		30	2100	1270	NA	1906.41	20
		40	2400	1270	NA	2372.37	0
40	20	20	1600	860	713.2	219.43	100
		30	2000	860	799.8	0.93	100
		40	2400	860	800	0.18	100
	30	20	2000	1270	NA	1913.46	20
		30	2400	1270	NA	1971.43	20
		40	2800	1270	1199.4	2.06	100
	40	20	2400	1680	NA	2366.15	0
		30	2800	1680	NA	2350.09	0
		40	3200	1680	NA	1896.05	20
Average			1800	960	NA	1036.9	59.26

Table 2: Performance of Cplex for the medium size *MPSC2* instances

$\lvert T\rvert$	$\lvert J\rvert$	$\lvert I\rvert$	$\lvert var\rvert$	$\lvert const\rvert$	z^*	cpu	$opt\%$
60	60	60	7200	3720	NA	2334.19	0
		70	7800	3720	NA	598.25	80
		80	8400	3720	3600	2.42	100
	70	60	7800	4330	NA	2041.46	0
		70	8400	4330	NA	2114.74	0
		80	9000	4330	NA	863.65	80
	80	60	8400	4940	NA	2295.5	0
		70	9000	4940	NA	2205.75	0
		80	9600	4940	NA	2125.92	0
70	60	60	8400	4330	NA	2219.95	0
		70	9100	4330	NA	1098	60
		80	9800	4330	4200	3.14	100
	70	60	9100	5040	NA	2108.07	0
		70	9800	5040	NA	2202.56	0
		80	10500	5040	NA	1638.3	40
	80	60	9800	5750	NA	2170.76	0
		70	10500	5750	NA	2102.91	0
		80	11200	5750	NA	2208.82	0
80	60	60	9600	4940	NA	2244.96	0
		70	10400	4940	NA	1888.41	20
		80	11200	4940	4800	4.16	100
	70	60	10400	5750	NA	2187.27	0
		70	11200	5750	NA	2193.29	0
		80	12000	5750	NA	1726.29	20
	80	60	11200	6560	NA	2184.66	0
		70	12000	6560	NA	2248.54	0
		80	12800	6560	NA	2112.58	0
Average			9800	5040	NA	1745.35	22.22

Authors Index