COMPLEX TASK TRAINING AND TACIT KNOWLEDGE TRANSFER IN TECHNOLOGY-ENABLED PEER ASSESSMENT ENVIRONMENT

Dmytro Babik, Hamid Nemati, Eric W. Ford
The University of North Carolina at Greensboro
Box 26170, Greensboro, NC 27402
(336) 334-5666
{d_babik, hrnemati, ewford}@uncg.edu

ABSTRACT

Despite popularity and validity of peer assessment techniques, several critical limitations are present in existing designs of IT-enabled peer assessment systems. These limitations are discussed and Double-Loop Mutual Assessment (DLMA) method is presented and justified as a novel technique for acquiring and assessing complex task competencies through social practice.

Keywords: Complex tasks, peer assessment, formative assessment, summative assessment, feedback

INTRODUCTION

Complex tasks are difficult to master. Simple tasks have a straightforward desired outcome, a single solution scheme, and no conflicting interdependence or solution scheme/outcome uncertainty (Zigurs & Buckland, 1998). In contrast, complex tasks are characterized by various combinations of complexity attributes, such as outcome multiplicity, solution scheme multiplicity, conflicting interdependence, and solution scheme/outcome uncertainty (Campbell, 1988). Writing essays, creating compositions and producing academic articles are but a few examples of complex tasks. Competency of dealing with complex tasks requires a combination of explicit and tacit knowledge. By definition, tacit knowledge, as opposed to explicit knowledge, is difficult to transfer from one person to another by means of formalization (Tsoukas, 2003). It can predominantly be acquired through the social practice of solving tasks and, therefore, resides in that practice (Cook & Brown, 1999; Hildreth & Kimble, 2002; Ribeiro & Collins, 2007; Tsoukas, 2003, as cited by Nonaka & Von Krogh, 2009). Two of the most important components of problem-solving practice are receiving feedback and integrating that feedback in future exercises. If mastering complex task requires a considerable transfer of tacit knowledge through practice, how can we systematically utilize feedback produced by social practice to help develop complex task competency? The overarching goal of this research is to develop a complex task competency-building system based on peer feedback.

The Socratic Method is a time-tested method of transferring tacit knowledge through deliberate practice and observational feedback. In an ideal world, the mentor poses a question to a group of learners who then debate amongst themselves on different ideas, solutions and their merits and impact. Then, rather than giving an answer, the mentor poses another question and learning goes on. In this idealized world, all instructors are sage but distanced, all students are engaged and positive. This method is extraordinarily effective because it pushes learners outside their comfort zone to explore alternative views and develop critical thinking. In the less than ideal world, however, the tradition has been to pass explicit and tacit knowledge from a mentor to the learner directly through lecturing and showing examples. Such is the conventional education system in most of the world today. This approach is less effective in transferring tacit knowledge and assessing complex task performance because it lacks practice opportunities and extensive feedback. Over the last thirty years, however, collaborative learning and peer assessment have gained popularity among educators as an approach to develop competencies of dealing with complex tasks and to assess performance (Topping, 1998).

In situations when assessment, or the success appraisal, of a task output cannot be attained objectively, peer assessment allows developing shared meaning of the value of a particular output. Peer assessment is an influential social learning strategy that allows subjects to become more involved in developing their competencies (Sluijsmans, Dochy, & Moerkerke, 1998). It is also a valid appraisal technique (Freeman, 1995; Lin, Liu, & Yuan, 2001). Peer assessment can be of two types – summative and formative. Summative assessment is intended to measure a subjects' attainment at a particular time, often for the purposes of external accountability (in the form of a score or a grade). Formative assessment, in contrast, is a set of formal and informal evaluation procedures with the express purpose of improving subjects' competencies through behavior modification (Croocks, 2001). Formative assessment involves qualitative feedback rather than quantitative scores (Huhta, 2008). Effective peer assessment typically requires close engagement of subjects into social interaction.

In certain environments, however, direct face-to-face interaction among peers is obstructed or even infeasible. Examples of such environments include large and/or asynchronous writing-intensive college courses and online collaboration (wikis, blogs, photography sharing). In other environments, such as academic journal submissions and reviews, and academic and corporate annual performance reviews, direct face-to-face interaction may be undesirable to assure anonymity and eliminate biases. Emulating peer assessment interactions in such environments is very difficult using traditional media. The proliferation of the modern paradigms of information system (IS) and telecommunications, such web 2.0, and the prevalence of creative content produce together an environment in which co-creation and collaborative learning face challenges and present opportunities unimaginable in traditional media environment. Modern information technology (IT) can be leveraged not only to reduce the resource load of peer assessment, but also enable stimuli, such as simultaneity, anonymity, random matching, reciprocal and pluralistic feedback. These stimuli may result in behaviors which are either too expensive or impossible to provoke otherwise. IT-enabled peer assessment, when embedded in e-learning or virtual co-creation environments, could increase interaction and communication among involved subjects, thus reducing feeling of isolation (Wegerif, 1998).

Despite their appeal in developing and assessing complex task competencies, many peer assessment systems proposed to date suffer severe deficiencies. They may excessively consume resources, create hostile environments, and be susceptible to assessment biases (Doiron, 2003; Topping, 1998; Trahasch, 2004). In peer assessment methods existing today, elimination of some of these deficiencies inevitably leads to an amplification of others. For instance, to control for biases the moderating role of central authority (such as instructor) is often advocated (Bostock, 2000), which, in turn, may lead to excessive use of instructor's time. Moreover, existing peer assessment systems are not comprehensive: they typically focus only on one of the two types of assessment (summative or formative) but not both. Arguably, these deficiencies can be eliminated by setting up peer assessment as a virtual social system that would enable close, yet anonymous, interaction of peers, the exchange of formative and summative feedback and generation of success metric. Such peer assessment system would present perhaps one of the most effective ways to produce desired outcomes – cultivate complex task competencies and generate valid and reliable success metrics. Comprehensive online peer assessment information systems (IS) suitable for use in various environments and social settings that satisfy these requirements, to the best of our knowledge, are not yet commercially available.

This study is motivated by the gap in the design of IT-enabled peer assessment suitable for developing complex task competency. We treat peer assessment as a social learning phenomenon. Our research questions are:

How peer assessment can be modeled as a decentralized and self-regulating social system?

and

How can such model be implemented as an IT-enabled social system for complex task assessment?

Self-regulating property of the system means that it should produce improved competencies, as well as valid and reliable success metrics. Decentralization property means that it should do so without direct intervention or moderation by a central authority, such as instructor or external expert. The purpose of the study is to present a new usable peer assessment artifact – the Double-Loop Mutual Assessment (DLMA) method. DLMA is designed to be applicable in a wide range of settings and embeds features that are expected to eliminate deficiencies of existing peer assessment systems.

This research makes the following contributions. Firstly, it contributes to the existing theoretical literature on peer assessment by introducing a novel peer assessment method coping with weaknesses of peer assessment methods designed and examined to date. Secondly, this research is of interest to designers of education IS because it presents a framework for developing a computational peer assessment algorithm. Finally, it opens a window of opportunity for practitioners to employ a novel IT-enabled peer assessment technique that offers substantial benefits in developing and assessing complex task competencies in a wide variety of settings.

RESEARCH DESIGN

Let's take a look at how the outcomes of DLMA could be examined. The design assumes a certain structure and relationships within the social system. A basic unit of interaction within DLMA is a dyad (Subject i – Subject j). The physical interaction within the dyad involves a sequence of anonymous reciprocal exchanges of (i) essays (representations of complex task solutions) and (ii) feedback of essays (formative assessment). In addition, each subject provides and the IS captures and stores (iii) summative assessment of other peer's essay according to certain criteria, (iv) summative assessment of other peer's feedback, again, according to certain criteria. These items are instances of information, perceptions and understanding that occur between the subjects in the dyad. Summative feedback is converted to a score according to a certain algorithm. A pool of subjects (e.g. a class of students) is divided into several peer groups of n subjects each; thus, each group consists of $n!/2(n-2)!$ dyads (e.g., a group of six students consists of 15 dyads) that engage in a virtually simultaneous interaction. Subjects are assigned to groups randomly. After each iteration (assignment) groups are re-matched. It is the ensemble of these dyadic interactions within a peer group (the DLMA treatment) that results in self-regulating learning and success metrics. We can hypothesize that interactions within this type of social system can (a) produce learning effect (in other words, a subject's competency shifts from initial (pre-treatment) level (g_0) to post-treatment level (g_K) by learning effect Δg); (b) result in success metric (cumulative score) that reflects the post-treatment level (g_K); and (c) eliminate the intervention of a regulating authority.

This study involves two units of analysis. At the process level, the unit of analysis is a peer group because it is the minimal level at which the DLMA method of assessment can be applied. At the analysis level, the unit of analysis is a subject because we are interested in the relationship of the distributions of subjects' competency and success metrics.

The DLMA workflow is executed according to the following protocol:

Rule 1: A pool of subjects is divided into peer groups of a sufficiently large size to assure validity but small enough to avoid cognitive overload.

Rule 2: Peer groups are composed (or re-composed) at random before each assignment.

Rule 3: In each assignment, all subjects in each peer group work on the same task (essays).

Rule 4: In each assignment, all subjects review each other's essays and receive textual feedback anonymously.

Rule 5: In each assignment, most subjects in each peer group submit their essays.

Rule 6: In each assignment, most subjects in each peer group submit their formative assessment (textual feedback) and summative assessment to their peers' essays.

Rule 7: In each assignment, most subjects in each peer group submit their summative assessment (feedback evaluation) of textual feedback received from their peers.

Rule 8: Sufficiently many assignments are given to assure validity but not too many to result in cognitive overload.

We hypothesize that under this protocol, DLMA produces two desired results. Firstly, over the course of several assignments, subjects exhibit a significant learning effect, that is, improved competency (e.g., creative and critical thinking) in performing a targeted type of complex task. Secondly, the summative assessment produced as a result of subjects' interaction reflects the distribution of the competency in a larger pool of subjects (for example, a class of students). Formally, this protocol should comply with the following measurement validity preconditions. Note that preconditions with the subindex ***a*** describe relationships in the DLMA system if the summative assessment mode based on ranking (that is, relative scale, or forced distribution) is selected, and propositions with the subindex ***b*** – if the summative assessment mode based on rating (i.e. absolute scale, or the Likert scale) is selected.

Precondition 1a: Under the rules 1 – 7, the observed within-group distribution of the average scores based on *ranking* summative assessment of essays approximates the latent distribution of the quality of essays within a peer group.

Precondition 1b: Under the rules 1 – 7, the observed within-group distribution of the average scores based on *rating* summative assessment of essays approximates the latent distribution of the quality of essays within a peer group.

Precondition 2a: Under the rules 1 – 7, the observed within-group distribution of the average scores based on *ranking* (relative-scale, or forced-distribution) summative assessment of *textual feedback* approximates the latent distribution of the quality of verbal feedback within a peer group.

Precondition 2b: Under the rules 1 – 7, the observed within-group distribution of the average scores based on *rating* (absolute-scale, or the Likert scale) summative assessment of *textual feedback* approximates the latent distribution of the quality of verbal feedback within a peer group.

Precondition 3a: Under the rules 1 – 7, the observed within-group distribution of the sum of the average scores for essay and verbal feedback based on *ranking* approximates the latent distribution of the current level of competency within a peer group.

Precondition 3b: Under the rules 1 – 7, the observed within-group distribution of the sum of the average scores for essay and verbal feedback based on *rating* approximates the latent distribution of the current level of competency within a peer group.

Precondition 4a: Under the rules 1 – 7, the observed pool-wide distribution of the sum of the average scores for essay and verbal feedback based on *ranking* approximates the latent distribution of the current level of competency in the pool of subjects.

Precondition 4b: Under the rules 1 – 7, for a given assignment, the observed pool-wide distribution of the sum of the average scores for essay and verbal feedback based on *rating* approximates the latent distribution of the *current* level of competency in the pool of subjects.

Precondition 5a: Under the rules 1 – 7, over a *series* of assignments, the observed pool-wide distribution of the *cumulative* sum of the average scores for essay and verbal feedback based on *ranking* approximates the latent distribution of the *terminal* level of competency in the pool of subjects.

Precondition 5b: Under the rules 1 – 7, over a *series* of assignments, the observed pool-wide distribution of the *cumulative* sum of the average scores for essay and verbal feedback based on *rating* approximates the latent distribution of the *terminal* level of competency in the pool of subjects.

The learning effect under DLMA treatment can be stated in the form of the following hypotheses (again, hypotheses with the subindex ***a*** apply to the summative assessment mode based on ranking, and hypotheses with the subindex ***b*** – if the summative assessment mode based on rating:

Hypothesis H1a: The *mean* of the pool-wide observed distribution of the *terminal ranking*-based *cumulative* scores (and, consequently, the latent distribution of the terminal level of competency) is greater than the *mean* of the pool-wide observed distribution of the *initial ranking*-based *cumulative* scores (and, consequently, the latent distribution of the initial level of competency).

Hypothesis H1b: The *mean* of the pool-wide observed distribution of the *terminal rating*-based *cumulative* scores (and, consequently, the latent distribution of the terminal level of competency) is greater than the *mean* of the pool-wide observed distribution of the *initial rating*-based *cumulative* scores (and, consequently, the latent distribution of the initial level of competency).

Hypothesis H2a: The *variance* of the pool-wide observed distribution of the *terminal ranking*-based *cumulative* scores (and, consequently, the latent distribution of the terminal level of competency) is *smaller* than the *variance* of the pool-wide observed distribution of the *initial ranking*-based *cumulative* scores (and, consequently, the latent distribution of the initial level of competency).

Hypothesis H2b: The *variance* of the pool-wide observed distribution of the *terminal rating*-based *cumulative* scores (and, consequently, the latent distribution of the terminal level of competency) is *smaller* than the *variance* of the pool-wide observed distribution of the *initial rating*-based *cumulative* scores (and, consequently, the latent distribution of the initial level of competency).

Note that hypotheses 1a, 1b, 2a and 2b describe the learning effect that entails the overall improvement of competency and the decrease in variance in the competency on the pool of subjects.

The experiment design would involve two randomly assigned pools of subjects: one be tutored in a traditional, instructor-to-student, manner (control group) and another receiving DLMA treatment as a means of facilitating collaborative learning and generating success metric. The latent construct of competency should be measured before and after the treatment in both pools and assured to have statistically similar distributions in the pools before treatment. Once treatments were administered, scores (observable construct measures) will be collected along with terminal (post-treatment) measures of latent competency. Hypotheses H1a, H2a, H1b, H2b can be falsified base on statistical testing of the differences in distribution parameters.

PRELIMINARY RESULTS AND CONCLUSIONS

Based on the premises of the Social Cognitive Theory and inviting potential of the web for collaborative learning, an IT-enabled domain-independent peer assessment system, named Double-Loop Mutual Assessment (DLMA) system, was developed. DLMA technique encompasses a simple but effective system of checks and balances that produces enhanced competency development and assessment through formative and summative feedback. The resulting success metric is valid, reliable and does not require intervention of a moderator.

We do not claim the DLMA approach to be a universal substitute for other learning and assessment approaches. We would like, however, to highlight the promise of this technique for the development and assessment of complex task competencies, such as creativity and critical thinking. At this time, only preliminary empirical results produced with the beta-version of DLMA are available. They support the presented propositions.

The further research on DLMA system will entail several new studies to test the presented propositions. Specifically, the following research questions still need to be addressed: 1) What are the effects, advantages and disadvantages of using alternative modes of summative assessment (ranking and rating)? 2) Is summative assessment produced by subjects (in each of these modes) valid vis-à-vis independent expert's assessment? 3) Is summative assessment inter-rater reliable? 4) What are subjects' attitudes toward DLMA and how do they relate to performance? Although there is yet to be much discovered about IT-enabled peer assessment, DLMA has a potential to deliver improved complex tasks' competency attainment in a wide variety of settings.

REFERENCES

Bostock, S. (2000). Student peer assessment. Centre for Learning Technology, Keele University.

Campbell, D. J. (1988). Task Complexity: A Review and Analysis. *Academy of Management Review*, *13*(1), 40–52.

Cook, S. D. N., & Brown, J. S. (1999). Bridging epistemologies: the generative dance between organizational knowledge and organizational knowing. *Organization Science*, *10*(4), 381–400.

Doiron, G. (2003a). The value of online student peer review, evaluation and feedback in higher education. *CDTL Brief*, *6*(9), 1–2.

Doiron, G. (2003b). The value of online student peer review, evaluation and feedback in higher education.

Freeman, M. (1995). Peer Assessment by Groups of Group Work. *Assessment & Evaluation in Higher Education*, *20*(3), 289–300.

Hildreth, P., & Kimble, C. (2002). The duality of knowledge. *Information Research*, *8*(1).

Huhta, A. (2008). Diagnostic and formative assessment. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics.*, Blackwell handbooks in linguistics (pp. 469–482). Malden: Blackwell Publishing.

Johnson, R. D., & Marakas, G. M. (2000). Research Report: The Role of Behavioral Modeling in Computer Skills Acquisition--Toward Refinement of the Model. *Information Systems Research*, *11*(4), 402.

Lin, S. S.-J., Liu, E. Z.-F., & Yuan, S.-M. (2001). Web-Based Peer Assessment: Does Attitude Influence Achievement? *IEEE TRANSACTIONS ON EDUCATION*, *44*, 211.

Nonaka, I., & Von Krogh, G. (2009). Tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory. *Organization Science*, *20*(3), 635–652.

Ribeiro, R., & Collins, H. (2007). The Bread-Making Machine: Tacit Knowledge and Two Types of Action. *ORGANIZATION STUDIES -BERLIN- EUROPEAN GROUP FOR ORGANIZATIONAL STUDIES-*, *28*(9), 1417–1434.

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1998). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, *1*(3), 293–319. doi:10.1023/A:1009932704458

Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, *68*(3), 249 –276. doi:10.3102/00346543068003249

Trahasch, S. (2004). From peer assessment towards collaborative learning. *Frontiers in Education, 2004. FIE 2004. 34th Annual* (p. F3F–16). Presented at the 34th ASEE/IEEE Frontiers in Education Conference, Savannah, GA.

Tsoukas, H. (2003). Do we Really Understand Tacit Knowledge? Blackwell.

Wegerif, R. (1998). The social dimension of asynchronous learning networks. *Journal of Asynchronous Learning Networks*, *2*(1).

Zigurs, I., & Buckland, B. K. (1998). A Theory of Task/Technology Fit and Group Support Systems Effectiveness. *MIS Quarterly*, *22*(3), 313–334.