Co-Creation and Iterative Social Learning in Technology-Enabled Peer Assessment Environment (Double-Loop Mutual Assessment Approach)

Research-in-Progress

Dmytro Babik
UNC - Greensboro
d_babik@uncg.edu

Eric W. Ford
UNC - Greensboro
ewford@uncg.edu

Abstract

This paper is motivated by IT-enabled peer assessment of complex task competencies as a social learning phenomenon. Despite popularity and evident validity of peer assessment techniques, several critical limitations, such as reliance on central authority and constraint to either formative or summative assessment, are present in existing designs of IT-enabled peer assessment systems. The present study contributes to the area of peer assessment literature by proposing theoretical justification for a new peer assessment technique, called Double-Loop Mutual Assessment (DLMA). This technique encompasses a simple but effective system of checks and balances that produces enhanced complex task competency development and assessment through formative and summative feedback. The resulting laissez-faire success metric is valid, reliable and does not require intervention of a moderator. The proposed technique can be applied in a wide variety of collective learning and/or co-creation setting where direct face-to-face interactions are difficult or anonymity is required.

Keywords

Social learning, peer assessment, critical thinking, double-loop learning, complex task competencies.

Introduction

Complex tasks are difficult to master. Simple tasks have a simple desired outcome, a single solution scheme, and no conflicting interdependence or solution scheme/outcome uncertainty (Zigurs and 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. Tacit knowledge, as opposed to formal or explicit knowledge, is difficult to transfer from one person to another by means of formalization. It can predominantly be acquired through the social practice of solving tasks and, therefore, resides in that practice (Cook and Brown 1999; Hildreth and Kimble 2002; Ribeiro and Collins 2007; Tsoukas 2003, as cited by Nonaka and Von Krogh 2009). Thus, mastering complex task requires a significant transfer of tacit knowledge through practice.

Over the last thirty years, peer assessment has gained popularity among educators as an approach to develop competencies of dealing with complex tasks and to assess performance (Topping 1998). In situations where assessment, or the success appraisal, of the output of a task 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 and Moerkerke 1998), and a valid appraisal technique (Freeman 1995; Lin, Liu and Yuan 2001). 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, 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 such peer assessment environment is very difficult using traditional media (“paper and pencil”). The proliferation of the modern information and communication technologies (ICT) and the prevalence of creative content together produce environment in which co-creation and collaborative learning face challenges and present opportunities unimaginable in traditional media environment. Modern ICT 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. All these stimuli allow behaviors which are too expensive or impossible to produce otherwise. IT-enabled peer assessment, when embedded in e-learning or virtual co-creation environments, would 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 create hostile environments, excessively consume resources (such as time or paper), 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 role of central authority (such as instructor) is often advocated (Bostock 2000), which, in turn, may lead to excessive use of instructor’s time. If these deficiencies were eliminated by the means of a virtual laissez-faire economy, a well-designed 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 the requirements of the posed research questions, to the best of our knowledge, are not yet commercially available.

This paper is motivated by IT-enabled peer assessment of complex task competencies as a social learning phenomenon. A principal research question is how an IT-enabled peer assessment system can be modeled and designed as a laissez-faire economy, so that it would produce sustainable equilibrium learning effect (improved competencies) and valid and reliable success metrics. The purpose of the paper is to present a new usable peer assessment artifact – Double-Loop Mutual Assessment (DLMA) method. An online project that instantiates the method has been launched in 2011 and is being tested in a south-eastern university.

This research carries implications for several audiences. Firstly, it offers a contribution to the existing theoretical literature on peer assessment by presenting a novel method that is expected to eliminate drawbacks of peer assessment methods designed and examined to date. Secondly, this research is of interest to designers of education information systems because it presents a framework for developing a computational peer assessment algorithm. Finally, it opens a window of opportunity for education 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. In the following sections, an overview of the theoretical foundations of DLMA is presented; various functions embedded in DLMA are explained and the set of assumption and propositions underlying DLMA functionality are formalized.

Theoretical Foundations of Collaborative Learning and Peer Assessment

Overview of Peer Assessment

To examine the premises and shortcomings of IT-enabled peer assessment and to review its current state, we focus on the intersection of two literature domains – peer assessment and computer-aided assessment. We leave outside the scope of this paper any types of computer-based testing. Although definitely having its merits, a multiple-choice test is not the best assessment tool for evaluating higher level skills of analysis, reasoning, synthesis and evaluation. We also leave out various approaches to self-assessment which are generally limited to a basic cognitive level (Anderson et al. 2001; Bloom, Krathwohl and Masia 1956). Peer assessment is aligned with two modern trends in education, namely competency-based
Iterative Social Learning in Double-Loop Mutual Assessment

Over the last three decades, researchers and instructors in social disciplines and education have been intensively looking into and experimenting with peer assessment of student performance (Topping 2005). In general, peer assessment refers to an approach that requires subjects to review and evaluate other subjects’ work and provide feedback. Topping (1998) presented a very extensive literature review on peer assessments and proposed topology of peer assessment techniques used in colleges and universities. He defined peer assessment as “an agreement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status” (Topping 1998, p. 250). We note that such agreement may be either explicit (in the case of open discussion) or implicit (in anonymous setting, such as double-blind review). Peer assessment in education has been used for many years (Boud, Cohen and Sampson 1999; Dochy, Segers and Sluijsmans 1999; Mowl and Pain 1995) and holds much promise as becoming a mainstream idea (Brown and Dove 1991).

While a wide range of research scenarios of peer assessment is presented in the literature, research on peer assessment could be broadly divided, by application context, into peer assessment of the individual performance and peer assessment of individual contributions in collaborative projects. Business schools in particular face challenges related to both types of peer assessment because they heavily rely, on the one hand, on case studies that require creative writing and, on the other hand, team projects that call for some measure of individual contributions (Brown, Rust and Gibbs 1994; Mello 1993).

By the purpose and content, peer assessment can be categorized either as formative or summative. Summative assessment seeks to monitor performance (Shepard 2007). In other words, summative assessment is intended to summarize the student’s attainment of a particular task at a particular time, often for purposes of external accountability (typically in the form of a score, grade, mark etc.). Summative peer assessment may be done using two different types of comparison – ranking or rating. These two terms sometimes are used interchangeably but there is a clear distinction between them (Vovici Blog 2011). Rating refers to the comparison of different items using a common (absolute) scale. Ranking means comparing different items directly one to another (on a relative scale). Ranking is sometimes called forced-distribution rating (e.g., in Schleicher, Bull and Green 2008). Both types of comparison – rating and ranking – have their strengths and weaknesses, and so far there has been no consensus in the literature as for which of them has a greater predictive validity (Krosnick 1999; Krosnick, Thomas and Shaeffer 2003). Formative assessment, in contrast, typically involves qualitative (verbal) feedback rather than quantitative scores (Huhta 2008). Formative assessment is a set of formal and informal assessment procedures employed by an instructor during the learning process with intent to promote further improvement of student attainment by modifying teaching and learning activities (Crooks 2001).

Collaborative Learning and Formative Peer Assessment

Feedback is an important aspect of learning that focuses on the details of content and performance. Moreover, it is an important feature of everyday social interactions outside the classroom and in the virtual space, and a core management competency to be mastered. Development of individuals’ growth in meta-cognitive and critical thinking skills can be enhanced by feedback (Wang and Wu 2002). Giving feedback and adjusting own behavior based on feedback is a skill that can, and should, be acquired through practice and training (Sluijsmans, Brand-Gruwel, van Merrienboer and Martens 2004). Peer assessment adds value as a learning tool by exposing students to the practice of evaluating others’ performance and receiving feedback on one’s own performance (Brutus and Donia 2010). Moreover, peer assessment enables learning at high cognitive levels (Bouzidi and Jailet 2009).

In view of the posed research question of how an IT-enabled peer assessment system can be modeled and designed as a laissez-faire economy, so that it would produce sustainable equilibrium learning effect and valid and reliable success metrics, we are interested in two desired outcomes – the positive effect on
learning and advanced creativity, as well as validity and reliability of success metric. The impact of peer assessment on learning and competency building is informed by theories in cognitive and social psychology domain (Yu, Liu and Chan 2005), specifically by the theory of information-processing (Gagné 1985; Reigeluth 1983; Wittrock 1978), the theory of social construction of knowledge (King 1989; Slavin 1992), and the theory of social modeling (Bandura 1985). In the light of information processing theory, practicing in performing complex tasks engages cognitive processes of organization, elaboration and rehearsal that advance cognitive development. According to the theory of social construction of knowledge, peer assessment is conducive to cognitive development because interaction among peers facilitates absorption of critical concepts (King 1989). When subjects interact with their peers in learning situations, disequilibration occurs, inconsistent knowledge is exposed, opposing perceptions and ideas are explored, and inadequate logical reasoning and strategies are challenged (Piaget and Gabain 1926; Slavin 1992; Yu et al. 2005). Intersubjectively derived common meaning enriches perspectives (Miranda and Saunders 2003) resulting in better comprehension by the subjects. Social modeling theory accounts for the positive effect of observing task performance by others on cognitive development (Bandura 1962; Bandura 1985). Social interaction presents opportunities for observing and imitating task performance by models. By allowing subjects to examine essays of their peers, as well as examine formative assessment (comments and suggestions) of their own work by the peers, they are exposed to exemplars for observation and modeling, which enhance subjects’ competency in the task. Through observing and modeling others performing the task successfully, learners can increase their self-efficacy, engage in the activity, gain experience and become interested in the task (Schunk 1995). Thus, the three presented theories together support the notion of subjects’ cognitive enhancement in the applied content domain through engagement in the process of creating, observing, and critiquing. Based on 109 empirical studies, Topping (1998) concluded that peer assessment produces positive formative effects on student achievement and attitudes. Specifically, he noted that these effects were “as good as or better than the effects of teacher assessment” (Topping 1998, p. 262).

**Summative Peer Assessment**

Validity and reliability of success metric produced by peer assessment techniques have been also supported empirically. A number of studies demonstrated that, generally, summative peer assessment is adequately valid and reliable in a wide variety of applications (Falchikov 1995; Falchikov and Goldfinch 2000; Topping 1998). This finding have been confirmed by more recent studies that used IT-enabled peer assessments (Cho, Schunn, and Wilson 2006; Zevenbergen 2001). Improved timeliness, frequency and volume of feedback can compensate lower individual assessment competency of subjects compared with expert assessment (Topping 1998). Despite that, Bouzidi and Jaillet (2009) called for further research on validity and reliability of peer assessment in a large study using a common metric and done across many courses and levels of students.

To summarize, peer assessment approach delivers numerous benefits to learning and co-creation for both the assessor and the assessee (Bostock 2000; Bouzidi and Jaillet 2009; Doiron 2003). Through social constructivism, peer assessment helps subjects recognize strengths and weaknesses of their creative process, identify target areas for remedial action, develop metacognitive skills and self-efficacy, and enhance their reflection and problem-solving abilities during collaborative learning. In addition, peer assessment enriches subjects’ interpersonal relationships (Sluijsmans et al. 2004).

Technology-enabled peer assessment information systems (IS) are capable of eliminating deficiencies of traditional-media-based peer assessment techniques (Lin, Liu, and Yuan 2001; Yu, Liu, and Chan 2005). They can facilitate more timely (sometimes instantaneous), convenient, individualized, anonymous feedback and permit global accessibility. IT-enabled peer assessment may allow monitoring of subjects’ progress at any point of assessment process. They are more manageable, less device-dependent, less resource consuming and more environmentally friendly than paper-and-pencil setups. In particular, IT-enabled peer assessment substantially reduces the instructor’s workload and the amount of paper necessary to facilitate assessment, especially in large classes (Rada, Michailidis and Wang 1994). For example, due to high student-to-instructor ratio in many education institutions, comprehensive writing tasks are either prohibitively effort-intensive or of a poor assessment quality and, consequently, are not used to full extent (Cho et al. 2006). With the increasing number of courses offered online with class sizes exceeding conventional face-to-face classes, quality grading of written assignments and providing timely feedback without peer assessment is becoming virtually impossible. While summative peer assessment is
generally related to reducing instructor’s workload, formative peer assessment is linked to enhancement of learning experience. However, there are also possible indirect cross-effects that need to be further investigated.

**IT-Enabled Peer Assessment**

IT-enabled (aka computer-aided, online or web-based) peer assessment emerged as a growing area in education in the late 1990s (Robinson 1999; Topping 1998). Historically, computer-aided assessment (CAA) has been primarily concerned with computer-based testing, and multiple-choice tests as a primary method because they relieved instructors from time-consuming and tedious process of grading. Only recently CAA has begun to include IT-enabled peer assessment. Topping (1998) suggested that IT-enabled peer assessment is a promising area for the future research and it has practical applications. The complexity of manual handling of the information flow of multiple peer assessment done repeatedly, especially in a double-blind format, with paper as a medium has been the main obstacle to wide use of peer assessment in the settings such as large academic classes. The use of modern IT can allow handling peer assessment in these settings with relative ease (Bostock 2000). IT-enabled peer assessment systems provide a means of more efficient assessment and, thus, offer a more suitable method for assessing the higher-level skills. The web has made collecting and redistributing feedback easy and cheap. What is necessary is an algorithm to handle formative and summative peer assessment efficiently, to create social structure that induces students to give honest and actionable feedback, and to output valid and reliable success metric.

**Designing New IT-Enabled Peer Assessment System**

Efficiency-related weaknesses of peer assessment, such as excessive consumption of resources or inability to accommodate efficient exchange of essays and feedback among a large number of subjects, can be successfully eliminated by designing an IS that would replicate the process in an IT-enabled environment. This has been successfully done in a number of applications (e.g. Brutus and Donia 2010; Trahasch 2004; Yu et al. 2005). Practically all of these applications support a very specific type of task, as well as either solely formative or summative assessment but not both. Moreover, most of these applications rely on an instructor as a guarantor of “fair game” and summative assessment validity.

The problem of how to stimulate subjects to give honest and actionable feedback and receive reliable scores without direct intervention of the instructor into the process of assessment has not been effectively addressed and presents an opportunity in peer assessment research. The present study contributes to theoretical literature and practice by proposing a novel peer assessment method. This method emulates a simple yet effective social system based on checks and balances, which induces subjects to develop creativity and critical thinking competencies and self-efficacy through formative assessment. Moreover, the system generates summative assessment as a result of subject-to-subject interaction without involvement of an instructor.

Bostock (2000) suggested that weaknesses of peer assessment may be tackled by anonymity, multiple assessors, and instructor moderation. While keeping the first two principles, we intended to substitute the hierarchical social system with the central role of an instructor as a moderator with a decentralized IT-enabled social system in which quality of formative feedback and validity and reliability of summative feedback are induced by reciprocity and pluralism. In addition, to attain positive effects of peer assessment, the number of rounds of peers assessment (or assignments) should sufficiently exceed two (Lin et al. 2001; Tseng and Tsai 2007). The proposed peer assessment systems, called Double-Loop Mutual Assessment (DLMA) system, is based on simultaneity and anonymity to encourage breadth of information sharing and frank communication in formative assessment (Miranda and Saunders 2003; Zhao 1998); reciprocity and pluralism to ensure validity and reliability of summative task assessment; and the double loop workflow, in which not only the essays but also formative feedback receive summative assessment that allows task closure (Miranda and Saunders 2003). Cho, Chung, King and Schunn (2008) empirically showed the efficacy of similar double-looped assessment approach in the non-expert knowledge refinement process. The conceptual model in Figure 1 positions DLMA with respect to other types of peer assessment approaches.
The DLMA workflow is structured in the following steps (the algebraic representation of the DLMA workflow will be presented in the paper submitted for journal publication):

Step 1: Subjects submit essays (representations of complex task solutions);

Step 2: Subjects rank-order (or rate) essays submitted in Step 1 by several of their peers and provide anonymous formative feedback to each of the peers in a reciprocal manner (that is, in a small peer group, everyone evaluates everyone else’s essay);

Step 3: Subjects rank-order formative feedback submitted in Step 2 in the same manner.

This workflow embeds the following basic conditions:

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

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

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

Condition 4: In each assignment, all subjects review each other’s essays and receive verbal feedback anonymously.

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

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

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

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

We assert that under these conditions, DLMA produces two desired results. Firstly, over the course of several assignments, subjects exhibit a significant learning effect, that is, improved creativity and critical thinking competency 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). More specifically, these desired outcomes can be formalized in the following set of propositions:

Proposition 1: Under conditions 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.

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

Proposition 3: Under conditions 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.

Proposition 4: Under conditions 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.

Proposition 5: Under conditions 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.

Proposition 6: Under conditions 1 – 8, 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).

Proposition 7: Under conditions 1 – 8, 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).

Note that propositions 6 and 7 describe the learning effect that encompasses the overall improvement of competency and the decrease in variance in the competency on the pool of subjects.

**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 *laissez-faire* 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? 5) does the distribution of summative assessment metric
resemble the distribution of competency? 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


Davies, P. 2001. “Computer Aided Assessment Must be More Than Multiple-Choice Tests for it to be Academically Credible?”


