THE EFFECT OF VISUALIZATION ON STUDENTS' MISCALIBRATION IN THE CONTEXT OF ONLINE PEER ASSESSMENT

ABSTRACT

Learning and course management systems (LMS/CMS), including the educational online peer review and assessment systems (EOPRAS), have been becoming more advanced and comprehensive in order to accomplish diverse pedagogical goals and accommodate various teaching and learning techniques. However, multidimensionality and complexity of information output produced by these systems creates barriers to understanding the results, and, consequently, impedes providing deep, unique and targeted feedback to learners. We see visualization as a possible solution to this problem. We describe a prototype visualization for complex peer assessment data and the empirical efficacy study.

Keywords: social learning, peer assessment, learning analytics, visualization, miscalibration.

INTRODUCTION

"...It's been a fine experience, but I feel as though I am not understanding where the overall grade for the assignment is coming from" wrote a student at a large public university in the US mid-Atlantic region in August 2016 answering a survey question about her experience with a particular online peer review system. Thanks to developments in information technology, the learning and course management systems (LMS/CMS), and the educational online peer review and assessment systems (EOPRAS) in particular, have been becoming more advanced and comprehensive in order to accomplish diverse pedagogical goals and accommodate various teaching and learning approaches, techniques and methods. Yet, multidimensionality and complexity of information output produced by these systems creates barriers to understanding the results for students and instructors, and, consequently, impedes providing deep, unique and targeted feedback to learners. While learning analytics do provide rich information about the learning process and outcomes, unless learners are able to genuinely understand this information, they cannot place their achievement in the context of the overall class' and their peers' performance and, therefore, cannot act upon it to improve their learning [6].

One way to lower the cognitive cost of understanding learning analytics is by providing intuitive visual representation, or *visualization*, of performance and learning outcomes data [12]. In general, "a picture is worth a thousand words" but, specifically, in the context of an LMS, intuitive visualizations offer the following benefits:

- Clarifying individual student performance on a task in relation to other peers and/or overall standing in the class while maintaining privacy;
- Gauging individual performance in more dimensions than a single reductionist assignment grade;
- Tracking trends of performance over time, including in comparison to peers;
- Visual overview of attainment of specific milestones and proficiencies.

As of now, unfortunately, most EOPRAS known to us do not offer these benefits because the state-of-the-art data visualization capabilities of the web environments, such as D3.js (`d3js.org`), FF Chartwell, Raphael.js, Dygraphs, have not been used in these systems.

The purpose of this paper is to succinctly describe a data visualization prototype specifically designed for EOPRAS and tentatively called the "*rainbow graph*", and to outline our empirical testing of its efficacy. We are interested in the effect of the visual representation of peer assessment results on student's performance miscalibration. *Miscalibration* refers to the dissimilarity of *self-assessment* (i.e., the learner's perception of their own attainment) and *external assessment* (i.e., other evaluators' perceptions of the learner's attainment) [9] [10] [16] [17]. In other words, miscalibration can be thought of as inaccuracy of self-assessment vis-a-vis aggregated multi-peer assessment. Following Kruger and Dunning (1999), if self-assessment exceeds aggregated multi-peer assessment, such miscalibration is called *overconfidence*. In contrast, if self-evaluation is lower than external evaluation, such miscalibration is referred to as *underconfidence*. Our research question is *Does data visualization of peer assessment interaction outcomes help learners understand their performance*?

Why is this important? Design and use of EOPRAS is a growing domain of pedagogical research and practice [3]. Systems, such as CPR, CritViz, Expertiza, Mobius SLIP, PeerCeptiv, are used in multiple higher education institutions to advance students' creative problem solving, critical thinking, communication and collaboration competencies [4] [5] [8] [15] [19]. Yet, professors and students regularly report the difficulty of understanding learning analytics and, therefore, using these systems [1]. The intricacy of outputs in the computerized peer review process requires design of novel, unconventional dynamic and interactive visual representations. Research and development efforts such as the PeerLogic project (`PeerLogic.org`) are aimed at creating and disseminating visualization web services and best practices for EOPRAS. This study contributes to the emerging research on the impact of learning analytics and data visualization in the context of LMS/CMS on advancing higher-level learning and cognitive skills such as creative problem solving, critical thinking, oral and written communication, team-work and collaboration.

BACKGROUND OF THIS STUDY AND RELATED RESEARCH

The scholarship of teaching and learning (SoTL) related to peer review and assessment has been evolving for forty years [20]. Multiple studies have evidenced its benefits in a variety of disciplines, e.g., accounting [13], engineering [11] [14], and social science [2] [7]. Peer assessment is aligned with two modern trends in education: competency-based curriculum, and the involvement of students in assessment. The overriding goal of peer assessment is to provide timely, rich and actionable feedback to learners. This feedback takes two forms: peers' verbal critiques and quantifiable evaluations. For any assessment to be of high quality it must be both accurate and effective [18], i.e., it must provide a valid representation of student achievement and actionable information to promote student learning.

Over the past twenty years, peer review has increasingly migrated to the IT-enabled online environment [21]. Online peer review and assessment process does not just mimic face-to-face interactions; it enables extended scope and scale of affordances. If implemented well, it facilitates and incentivizes learners to provide more extensive, detailed, reflective and critical feedback than typically can be done in oral or paper-and-pencil settings. Through peer feedback, the authors experience multiple perspectives on their work rather than the singular voice of a teacher. For the instructor, it generates multiple and diverse performance measures that can be used to judge the individual and overall class's performance and progress. These data can be also used to provide unique targeted feedback and even suggest grades for students.

Classrooms are complex cognitive domains where both learners and instructors can vary in skill level, are immersed in networks of social relationships, and where everyone is trying to navigate learning new skills and competencies, and to gain deeper understanding through trial and error. As peer review systems expand their capabilities, we are presented with a challenge and an opportunity to create new ways to visually depict these complex relationships, and help participants understand and influence these processes of creating meaning. We can build visual representations that establish strong correspondences between the peer review processes and the cognitive work of the learner.

Visualization of data generated by EOPRAS affords us the opportunity to transform online peer review in two ways. First, digital technologies allow us to build new visual representations of peer review processes and outcomes that can better depict cognitive work, relationships, trends and activities of the learners. Second, these representations can be dynamic, interactive and real-time, thus influencing the dynamic unfolding of the learner behavior.

As peer review takes these new forms thanks to EOPRAS, its purpose shifts from being primarily a method for generating ample peer feedback toward a way for the class to amass its own cognitive surplus and to rapidly direct the attention of the entire class to notable exemplars. These exemplars, both high and low quality works, can act as collective cognition markers or goal-posts that each individual learner can use to self-calibrate their own effort and performance level.

This comes to the question of how to build meaningful representations that expand EOPRAS' capabilities. We need to understand the design task as one of establishing a correspondence between the processes being represented (peer review) and the visual representation (interface) in the larger context of learners' and instructors' goals.

VISUALIZATION PROTOTYPE DESIGN

To help students understand their performance in the online peer review process, a group of researchers at Arizona State University designed a prototype of visual data representation tentatively called the "rainbow graph" (Figure 1). This dynamic, interactive graph is based on the open D3 JavaScript library (`d3js.org`) but is specifically engineered for visualizing multifaceted results of the peer assessment interactions in EOPRAS. The "rainbow graph" displays data from a single-round multi-peer assessment on the ordinal (ranking) or cardinal (rating) scales in a group of students (e.g., the whole class or a peer review group). Each column represents a recipient of evaluation, i.e., a student, whose submission was assessed by peers (other students). Depending on whether the graph is displayed in the instructor or student console, columns can be either identified by student names or displayed unlabeled (anonymized). When this graph is displayed in the student console, only the signed-in student's column is identified by their name; columns corresponding to students who provided evaluations are indicated, but names of peer reviewers are not shown. The overall height of the column indicates the submission attainment level as the score based on the aggregation (average, median, etc.) of scores given to the submission by multiple peer reviewers. The higher the column, the higher is attainment. Columns are always sorted by height (the higher to the left, the lower to the right), thus indicating standing of each student's submission among the rest of submissions in the class. The number of segments in each column indicates the number of peer reviewers who evaluated the submission. Finally, colors indicate individual scores given by peers, e.g., the shades of green indicate higher individual scores and the shades of orange indicate lower scores. The greener the segments in the column, the higher is the column. When a particular segment of a column is "hovered over", the graph interactively indicates the peer reviewer who provided that specific evaluation. The graph supports additional interactive elements to improve ease of use, such as selection of different color schemes (to accommodate needs of colorblind users) and scrolling for navigating the graph for a large class. The documentation and mock-ups of this visualization are available at `www.peerlogic.org/visualization-rainbowgraph`.

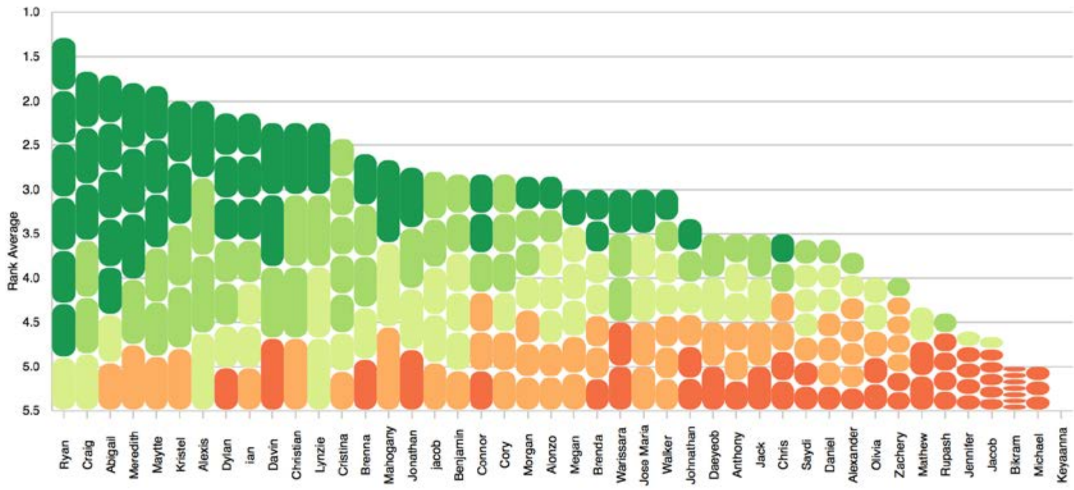

Figure 1 - The Dynamic Interactive Rainbow Graph

This visualization helps students and instructor understand the following aspects of peer review interaction and performance:

- The number of evaluations given to each submission and individual scores given by peer reviewers;
- Individual student's submission attainment (standing) in comparison with other peers' submission within a peer review groups and/or "globally" across the entire class;
- The degree to which peer reviewers' evaluations of a particular submission are similar of dissimilar (the variance or "controversy" of a submission);
- Submission attainment (standing) of submission made by peer reviewers who evaluated a particular submission.

The combination and interaction of these pieces of information presented in a single dynamic graph eases understanding on multiple complex facets of peer review outcomes. We believe it helps reduce the number of questions frequently asked by students who seek to understand these outcomes. The goal of this study is to evaluate the effectiveness of the "rainbow graph" in reducing students' miscalibration of their performance in peer review process. We hypothesize that the visual presentation of performance data via the "rainbow graph" as a complement to the numerical performance data presented in the tabulated format reduces miscalibration.

METHODOLOGY

In this study, we seek to examine the relationship between performance miscalibration, as the dependent variable, and the use of the "rainbow graph, as the treatment, or independent, variable. Therefore, we use the classic controlled experiment design (the randomized control-group pre-posttest design) to test our hypothesis that the presentation of multidimensional performance data as the "rainbow graph" (in addition to the standard interface with tabulated numerical performance data) has a significant negative effect on miscalibration.

The experiment will involve a class of at least 60 students taking a course requiring peer review and assessment of individual solutions to complex, open-ended problems, such as creating a visual art piece, writing an essay, developing a systems analysis diagram or a business process model. Since the principal investigators of this study have access to several different EOPRAS, we intend to replicate this study in several different classes of the same course using different systems in order to test the efficacy of the "rainbow graph" in different systems. In addition, we intend to replicate this study in undergraduate and graduate courses to be able to control for student level. In every class, students will be randomly assigned to the control or treatment groups of approximately equal size. Thus, we aim at collecting an overall sample of several hundred subjects, with the possibility of controlling for the student level, the course topic, and the EOPRAS used.

Miscalibration (the dependent variable) is measured as the difference between a student's self-assessment of their performance in the assignment and the aggregate peer evaluation of that student's attainment. Thus, positive miscalibration represents overconfidence and negative miscalibration represents underconfidence. The aggregate peer evaluation is measured as the average ranking or rating given by several peer reviewers to each student's submission on a particular scale, depending on the EOPRAS used. Respectively, student's self-evaluation will be elicited on the same scale. As an additional measure, z-score and percentile rank will be computed to capture the "global" standing of the student's submission among all submissions in the class. Respectively, each student will have to report their perceived z-score and/or percent rank (e.g., "top 1/3", "middle 1/3", "bottom 1/3" or a more precise scale). Students in the control group will be asked to self-assess, then presented the tabulated numeric performance data using the standard interface of the respective EOPRAS, and then asked to self-assess again. Students in the treatment group will be asked to self-assess, then presented the tabulated numeric performance data using the standard interface of the respective EOPRAS and the "rainbow graph", and then asked to self-assess again.

Data analysis will be conducted using the analysis of variance (ANOVA).

EXPECTED RESULTS AND FUTURE DIRECTIONS

The findings of this study will guide our further exploration of ways to improve design of EOPRAS interfaces and enhance delivery of learning analytics from IT-enabled social learning environments to students and instructors.

ACKNOWLEDGEMENT

This research is a part of the PeerLogic project (`PeerLogic.org`) sponsored by the NSF IUSE grants DUE-1431856, DUE-1431975, DUE- 1432347. We would like to acknowledge substantial contributions to this study made by Jaharsh Samayan Kotha Venkata Sadha and Rajat Aggarwal.

REFERENCES

[1] Alliance for Excellent Education, Capacity enablers and barriers for learning analytics: Implications for policy and practice. Report, June 2014.

[2] Althauser, R. & Darnall, K. Enhancing critical reading and writing through peer reviews: An exploration of assisted performance. *Teaching Sociology*, 2001, 29(1), 23–35.

[3] Babik, D., Gehringer, E., Kidd, J., Pramudianto, F. & Tinapple, D. Probing the landscape: Toward a systematic taxonomy of online peer assessment systems in education. In *CSPRED 2016: Workshop on Computer-Supported Peer Review in Education*, Raleigh, NC.

[4] Babik, D., Singh, R., Zhao, X. & Ford, E. What you think and what I think: Studying intersubjectivity in knowledge artifacts evaluation. *Information Systems Frontiers*, 2015, 1–26.

[5] Cho, K. & Schunn, C. D. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education*, 2007, 48(3), 409–426.

[6] Dietz-Uhler, B. & Hurn, J. E. Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, 2013, 12(1), 17–26.

[7] Falchikov, N. Peer feedback marking: Developing peer assessment. *Innovations in Education Training International*, 1995, 32(2), 175–187.
[8] Gehringer, E., Ehresman, L. M. & Skrien, D. J. Expertiza: Students helping to write an OOD text. In *Companion to the 21st ACM SIGPLAN Symposium on Object-oriented Programming Systems, Languages, and Applications*, New York, NY, USA, 2006, 901–906.
[9] Kruger, J. & Dunning, D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 1999, 77(6), 1121–1134.
[10] Kruger, J. & Dunning, D. Unskilled and unaware – but why? A reply to Krueger and Mueller. *Journal of Personality and Social Psychology*, 2002, 82(2), 189–192.
[11] MacAlpine, J. M. K. Improving and encouraging peer assessment of student presentations. *Assessment & Evaluation in Higher Education*, 1999, 24(1), 15–25.
[12] Olmos, M. M. & Corrin, L. Learning analytics: A case study of the process of design of visualizations. *Journal of Asynchronous Learning Networks*, 2012, 16(3), 39–49.
[13] Persons, O. S. Factors influencing students' peer evaluation in cooperative learning. *Journal of Education for Business*, 1998, 73(4), 225–229.
[14] Rafiq, Y. & Fullerton, H. Peer assessment of group projects in civil engineering. *Assessment & Evaluation in Higher Education*, 1996, 21(1), 69–81.
[15] Russell, A. A. Calibrated Peer Review: A writing and critical-thinking instructional tool. *UCLA, Chemistry*, 2001.
[16] Sadler, P. M. & Good, E. The impact of self-and peer-grading on student learning. *Educational Assessment*, 2006, 11(1), 1–31.
[17] Sargeant, J., Mann, K., van der Vleuten, C. & Metsemakers, J. 'Directed' self-assessment: Practice and feedback within a social context. *Journal of Continuing Education in the Health Professions*, 2008, 28(1), 47–54.
[18] Stiggins, R. J., Arter, J. A., Chappuis, J. & Chappuis, S. *Classroom Assessment for Student Learning: Doing it Right – Using it Well*. Assessment Training Institute, 2004.
[19] Tinapple, D., Olson, L. & Sadauskas, J. CritViz: Web-based software supporting peer critique in large creative classrooms. *Bulletin of the IEEE Technical Committee on Learning Technology*, 2013, 15(1), 29.
[20] Topping, K. J. Peer assessment between students in colleges and universities. *Review of Educational Research*, 1998, 68(3), 249–276.
[21] Topping, K. J. Trends in peer learning. *Educational Psychology*, 2005, 25(6), 631–645.