What do educational data, generated by an online platform, tell us about reciprocal web-based peer assessment?

Olia Tsivitanidou[1] and Andri Ioannou[1,2]

[1] Research center on Interactive media, Smart systems and Emerging technologies (RISE), Nicosia, Cyprus [2] Cyprus Interaction Lab, Cyprus University of Technology, Limassol, Cyprus
o.tsivitanidou@rise.org.cy, andri.i.ioannou@cut.ac.cy

Abstract. Peer Assessment (PA) is a promising evaluation strategy in the educational context, not only due to its effectiveness to reduce instructor's evaluation loading, but mainly due to its benefit towards student development e.g., teamwork, in-depth thinking. In this exploratory study we sought to explore how do educational data, as generated by an online platform (i.e., Peergrade) and displayed in teacher's and students' Learning Analytics Dashboard (LAD), can potentially inform us of the PA process and the peer interactions, as they take place. Participants in the study were 21 undergraduate teacher-students who attended a science course (electrical circuits topic) following the inquiry-based approach. Students were asked to reciprocally and individually assess the responses of a peer in a given task. The findings of this study have implications towards the establishment of new theoretical frameworks and developments for bridging educational theory, design process and data science, in the field of assessment.

Keywords: Web-based Peer Assessment, Peer Feedback, Online Assessment, Science Education, Science Learning.

1 Introduction

Peer Assessment (PA) constitutes an educational activity in which students judge the performance of their peers by offering oral and/or written peer feedback. It is often integrated in the wider context of formative assessment and it endorses many benefits in terms of students' learning especially when it is reciprocally implemented [4]. When employed formatively, PA can improve students' learning accomplishments [2] and their overall performance (e.g., specific skills and practices) in various domains including science education [3, 4]. PA has received attention in participative inquiry-oriented science learning environments, especially computer-supported learning environments and in recent years in Massive Online Open Courses (MOOCs) [1]. Yet, research on how educational data, as generated by online platforms and communicated via a Learning Analytics Dashboard (LAD) to teachers and students, can potentially inform us of the PA process and the interactions that take place among peers, is still scarce. The analysis and interpretation of such data can provide some insights into the emerging

field of Learning Analytics (LA), which has become a must in education, with the critical goal to use them for understanding and supporting learning. Among the challenges that still exist in this research area, is how to ensure the quality, timing, and form of feedback, which is critical to effective learning.

2 Rationale and Research Questions

We sought to explore how educational data, generated by an online platform and communicated via a LAD to teachers and students, can potentially inform us of the PA process and the interactions that take place among peers. We focused on data connected to both the role of the assessor and the assessee, so as to explore both segments of reciprocal peer assessment. Overall, we aimed to examine how LA data conjecturing peer feedback validity are presented via LAD and whether it can be interpreted in a meaningful manner. In view of the above, the following Research Questions (RQs) were sought to be addressed in this study: **RQ1:** How does the 'submission score' generated by the online platform used in this study, associate with the quality and validity of students' assessed artifacts (submissions)? **RQ2:** How is the median time spent per review per peer assessor, associated with the length of the qualitative peer feedback? **RQ3:** Did students, as assessees, proceed with revising their initial responses after the completion of PA? (If yes, did those revisions contribute towards improving the validity of their initial responses?) **RQ4:** How do students, as assessees, react to the peer feedback received? (Do reaction 'likes' relate to the reaction comments that follow and how?)

3 Methodology

The sample was consisted by 21 undergraduate teacher students (19 females and 2 males), who worked in groups of two while studying the learning material on the topic of Electric Circuits in the module of Electromagnetism of the Physics by Inquiry material following the inquiry-based approach. In a carefully chosen check point of the learning material, the students were asked to complete, on an individual basis, a diagnostic task comprised by three distinct questions e.g., *"List the light bulbs numbered as 1, 2 and 3 in the given circuit (figure provided), in a decreasing order of brightness. Explain your reasoning".* Upon completing the task and having submitted their responses in the Peergrade online platform (https://app.peergrade.io/), they implemented reciprocal web-based PA. The platform automatically assigned students' responses to their peers for peer evaluation. Students were asked to provide feedback to two peer submissions, via the 'review' tab of the platform, with the assistance of a given rubric, that was comprised by 3-point Likert scaled 10 assessment criteria. Students, as assessors, rated their peers' responses on 10 criteria e.g., *"The order in which the bulbs are classified in order of decreasing brightness is justified"*, in accordance with a 3-point Likert scale and also provided written comments for justifying their ratings. Each student individually assessed the responses of two other students which were automatically assigned to her/him. This review task lasted on average 20 minutes per evaluation,

with a quite substantial standard deviation in time (M = 20.3 min, SD = 7.8 min). After the implementation of the PA, students, as peer-assessees, were allowed to revise their responses, after studying the peer feedback comments that they received from two other peers, via the 'react' tab of the platform. This tab allows to students to react to peer feedback in the following manner: (a) like a feedback comment, (b) comment on a feedback comment, (c) flag an issue (merely teachers get informed about flags). At the end of the 'react' phase, students were asked to evaluate the peer feedback, by rating a 4-point Likert scale question (mandatory) and provide a comment (optional).

Data were collected from four sources; namely: (a) pre-instructional questionnaire; (b) data displayed in the LAD of the Peergrade online platform, i.e., the median time spent per review, per student, measured in minutes; the average word count of peer feedback comments per student; submission score generated upon completion of the peer review phase (the submission score was generated based on the scores that peers provided to the assessment criteria, while peer reviewing); feedback score generated for every student for each assignment (the feedback score for a reviewer is based on the feedback reactions s/he receives); assessees' reaction to peer feedback (likes and reaction comments); (c) students' initial and revised responses to a given diagnostic task; (d) audio recordings data resulted from Think Aloud Protocols (TAP), which were used during the provision and review of peer feedback by students, for triangulation purposes. A mixed-method approach was used that involved both qualitative and quantitative data.

4 Results

With respect to RQ1 we ran linear Pearson's r correlation to check the existence of potential correlation between submission scores (M = 0.64, SD = 0.10) that the Peergrade tool generates and the quality of initial responses (M = 2.80, SD = 2.27) to the given task, which resulted through open coding by the authors. The results indicate that there is no correlation between the two aforementioned variables $r = 0.160$, $p = > .05$, $n = 21$. We further explored how the median time spent per review per student (assessor) is associated with the length of the qualitative peer feedback (measured via word counts) for addressing RQ2. A positive correlation was found to exist between median time spent (M = 20.32 minutes, SD = 7.84) and the average word count of peer feedback comments per student (M = 341.8, SD = 147.4), Pearson's $r = 0.560$, $p = < .001$, $n = 21$. Think aloud protocols data shed light into the reasons behindhand difference in time spent for giving feedback among peer assessors (e.g., students, who spent more time while giving feedback, where those who were crosschecking their own responses, before providing feedback, and this additional activity implied more time needed).

In relation to the peer assessee role, findings of RQ3 revealed that out of the 21 students who received peer feedback in this study, 12 students (in the assessee role), proceeded with revising their responses. Wilcoxon Signed-Ranks Test outcomes indicated that the median post-test ranks were statistically significantly higher than the median pre-test ranks $Z = -2.132$, $p < 0.033$. With respect to 'reaction' data generated by the Peergrade tool (RQ4), a total of 403 valid entries for reaction likes and reaction

comments (as resulted from peer feedback data from all students) were provided. A negative correlation between reaction likes (M = 0.35, SD = 0.48) and reaction comments (M = 0.17, SD = 0.38), r = -0.248, p = < .001, n = 403 was found to exist. The qualitative analysis of the reaction comments, has shown that whenever students offered comments, instead of likes, that was mostly due to disagreements they had with their peers' feedback comments. This finding indicates that reaction likes can be interpreted in a meaningful manner, as they can provide signs to teachers on whether assessors and assessees agree or disagree on the peer feedback exchanged. Nevertheless, such a trend should be treated with caution, since disagreements in reaction comments offered by the assessees were identified in two different cases: (a) assessees disagreed with the content of the critical peer feedback received and insisted on the validity of their own initial response; (b) assessees disagreed with the content of the embracing peer feedback received and scrutinized the validity of their own initial response.

An in-depth analysis of the data is being conducted to answer the research questions of the study; we hope to present some of this further analysis during the conference. Overall, the proposed work is expected to have immediate implications in science teaching and learning but is also expected to inform formative assessment research and practice in different domains and contexts, namely in peer-assessment in blended, online learning and MOOC courses. Designers of web-based learning platforms and technological tools for education could utilize this piece of information in several manners already explicated above, e.g., framing and interpreting educational data for learning analytics derived from peer-assessment activities; developing appropriate tools for peer assessment.

Acknowledgements

This project has received funding for the European Union's Horizon 2020 research and innovation programme under grant agreement No 739578 and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

References

1. Alcarria, R., Bordel, B., de Andrs, D. M., & Robles, T.: Enhanced Peer Assessment in MOOC Evaluation Through Assignment and Review Analysis. International Journal of Emerging Technologies in Learning (iJET), 13(1), 206-219 (2018).
2. Falchikov, N.: Involving students in assessment. Psychology Learning and Teaching, 3, 102–108 (2003).
3. Tsivitanidou, O., & Constantinou, C.: A study of students' heuristics and strategy patterns in web-based reciprocal peer assessment for science learning. The Internet and Higher Education. 12, 12-22 (2016).
4. Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M.: Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. European Journal of Psychology of Education, 33(1), 51-73 (2018).