Online Peer Assessment: Comparing between In-class and Massive Open Online Course Students

Maya Usher

Technion – Israel Institute of Technology

ushermaya@tx.technion.ac.il

Miri Barak

Technion – Israel Institute of Technology

bmiriam@technion.ac.il

Abstract

The importance of peer assessment has increased with the rise of massive open online courses (MOOCs), as it provides an efficient way for grading hundreds of open assignments, when the number of instructors is limited. Yet, research on the quality of online peer assessment (grading and feedback) has not received sufficient attention. Hence, this exploratory case-study was set to examine the mechanism of online peer assessment by comparing two groups of students: In-class and MOOC. The first group studied a frontal course and assessed the work of their classmates, while the second group studied a MOOC and assessed the work of unfamiliar peers from different countries. Both groups studied the same learning materials with the same teaching team, and provided online peer assessment. Findings indicated that the in-class students awarded to each other significantly lower grades compared to the MOOC students. However, their grades were better correlated to those given by the teaching assistants. Content analysis of students' peer feedback indicated four categories, classified in an increasing order of quality: reinforcement, clarification, verification, and elaboration. The peer feedback of the in-class students was more associated with higher order thinking, compared to their MOOC counterparts.

Keywords: Engineering education, Higher education, Massive open online course, Online peer assessment, Peer feedback.

Introduction

Peer assessment is a pedagogical method generally referring to a process wherein students take part in judging the quality of their colleagues' learning outcomes (Sadler & Good, 2006; Topping, 1998). It is usually conducted in anonymity, supported by a ridged scoring rubric and a set of detailed instructions (Barak & Rafaeli, 2004; Luaces et al., 2015). A continuously growing body of research underlines the value of peer assessment for the learning process (Falchikov & Goldfinch, 2000). It was documented as a method for increasing motivation (Hanrahan & Isaacs, 2001), gaining positive effects on students achievements (Barak & Rafaeli, 2004), and providing a glimpse to the work from an assessor's perspective (Kulkarni et al., 2013). The process of evaluating the work of fellow learners may result in affective changes, building a greater sense of shared ownership for the learning process (Sadler & Good, 2006).

However, there is a dispute among educational researchers regarding the reliability and validity of peer assessment. Some studies found positive correlations between the grades students awarded their colleagues and those given by the teacher (Sadler & Good, 2006; Topping, 1998) or by the teaching assistants (Kulkarni et al., 2013). Other studies have reported several concerns led by questions of reliability, validity, and students' bias (Falchikov & Goldfinch, 2000). Kulkarni and colleagues (2014) have argued that students may hold systematic cognitive biases while assessing. Studies have also payed attention to gender or nationalistic stereotyping that might impede the objectivity of the evaluation (Kulkarni et al., 2013).

Online peer assessment was referenced to as a key feature that enables large online classes to use open assignments (Kulkarni et al., 2014). Its relevance intensifies especially in the context of massive open online courses (MOOCs), since they attract thousands of students from hundreds of countries simultaneously (Barak, Watted & Haick, 2016). In order for learning to be meaningful, and involve knowledge sharing, and social interactions, it is recommended to use open assignments (Çevik, 2015). However, evaluating open assignments in MOOCs is a difficult task because of the number of students involved (Luaces et al., 2015). Since it is unpractical that a small group of TAs will evaluate hundreds of open assignments, researchers have been exploring the possibilities of introducing students to the online peer assessment approach (Kulkarni et al., 2013).

Past studies focused either on peer assessment in the traditional classroom (Falchikov, 1995; Kollar & Fischer, 2010; Sadler & Good, 2006) or in online learning environments (Çevik, 2015; Cheng, Liang & Tsai, 2015; Gielen & De Wever, 2015). However, literature lacks studies that compare between the two different learning environments. In light of the aforesaid, this exploratory case-study aimed at understanding the mechanism of online peer assessment (numerical grading and written feedback), while comparing between in-class and MOOC students. This goal raised the following research questions:

(1) Are there differences and/or correlations in peer grading between in-class and MOOC students, and the TAs?

(2) Are there differences in peer feedback between in-class and MOOC students?

Methodology

Participants and procedures

This study included 53 undergraduate students in science and engineering, who enrolled to the same university and studied the same learning materials in Nanotechnology and Nanosensors in two courses: one delivered in a traditional face-to-face mode and the other online via MOOC. Hence, the study included two research groups: a. In-class students (N = 17), who assessed the work of their classmates, and b. MOOC students (N = 36), who assessed the work of unfamiliar peers from different countries. Both groups studied with the same teaching staff, and had identical assignments.

As part of the course requirements, the students were asked to work on an open assignment- to provide an example for an innovative nanoscale sensor they would like to use in their everyday life. After submitting their assignments, each student was randomly and anonymously assigned to assess the work of at least two peers. The in-class students used Moodle 'Workshop' tool; while the MOOC students used Coursera platform (www.coursera.org). Since the students in both groups were officially enrolled to the university, their work was also assessed by two teaching assistants (TAs). Both the TAs and students used the same scoring rubric that included seven grading criteria, on a scale of 0 to 5, with a maximum of 35 points that were normalized to 100. In addition to numerical grading, the students were encouraged to provide written comments that would help their peers improve their work.

Research method and data analysis

This study was based on the mixed methods research (Kellogg, Booth & Oliver, 2014); wherein, peer grading was analyzed quantitatively while peer feedback was examined qualitatively. *Peer grading* was examined by analyzing differences between the numerical scores of the two research groups and correlations between their grades and those of the TAs. Due to the small number of participants and the fact that normal distribution could not be assumed, we used nonparametric tests. Mann-Whitney test examined the differences between the two independent groups and Spearman correlations examined the strength and direction of association (Corder, & Foreman, 2014).

Peer feedback was examined by content analyzing students' written comments according to four coding categories, partially adapted from Gielen and De Wever (2015): Reinforcement, Clarification, Verification, and Elaboration, with an increasing level of feedback quality and thinking skills. The *Reinforcement* category refers to general comments with little, if at all, cognitive contribution for improvement. It includes two sub categories: positive and negative. The *Clarification* category refers to comments indicating lack of clarity. It includes two sub categories: Technical writing, and scientific knowledge. Both of those categories require little cognitive effort on behalf of the assessor, and therefore are conceptualized as lower order thinking skills. The *Verification* category refers to whether the work has met the requirements of the assignment, and includes two sub categories: validate and invalidate. The *Elaboration* category refers to comments assisting with relevant information to help the learner in error correction. It holds a distinction between informative elaborations (providing additional information regarding certain aspects of the assignment), and suggestive elaborations (providing ideas for improvement by identifying gaps in the work, and suggesting alternative solutions). *Elaboration* comments are conceptualized as the highest order thinking skills. These skills correspond with Bloom's Taxonomy of educational objectives (Bloom et al., 1956).

Findings

Peer grading: Comparing between in-class and MOOC students, and TAs

Findings indicated that the in-class students were inclined to award lower grades to their peers (M = 85.29, SD = 9.29) compared to the MOOC students (M = 93.47, SD = 9.99). Mann-Whitney test indicated that this difference was statistically significant (U = 134.00, p = .001). The peer grades of both groups were examined against the grades given by the TAs. Findings indicated that the in-class students awarded lower mean grades than those given by the TAs (M = 85.29, SD = 9.29; M = 90.12, SD = 5.1, respectively); while the MOOC students awarded similar mean grades to those of the TAs (M = 93.06, SD = 7.40; M = 93.47, SD = 9.99, respectively). These results might suggest that MOOC students were more "accurate" in providing grades to their peers; however this was not the case. According to Spearman test, only the grades of the in-class students were statistically positively correlated to those of the TAs (rs(17) = .48, p = .05). Our results suggest that the in-class students are more stringent in grading their peer, but their grades are in good correlation with those given by the TAs.

Peer feedback: Comparing between in-class and MOOC students

Overall, 17 (89%) of the in-class students and 36 (54.5%) of the MOOC students provided written comments, which were divided into short segments. Our deductive content analysis identified 116 comment segments for the in-class students and 283 comment segments for the MOOC students. In average, a slight advantage was recorded for the MOOC students in the number of segments per peer evaluation (M = 7.87, SD= 3.04), compared to the in-class students (M = 6.82, SD = 3.9). Table 1 represents examples of short segments for each category, by research group.

Table 1. Coding scheme for the content analysis of peer feedback messages

Category	Sub category	In-class Examples	MOOC Examples
1. Reinforcement	1.1 Positive	*Well done!*	*Great idea!*
	1.2 Negative	*Your article is a little long.*	*Your answer is closer to a joke.*
2. Clarification	2.1 Technical	*The English is poor.*	*It was hard to read and understand the context.*
	2.2 Scientific	________________	*I had a problem in understanding exactly what does the nanosensor sense and how it responds.*
3. Verification	3.1 Validate	*The idea is creative and relevant for everyday life.*	*The answer is based on concepts from the course and it is based on relevant sources.*
	3.2 Invalidate	*The element of creativity is missing … cannot be termed a new idea.*	*There was no innovative idea for a nanosensor in all 8 (!!!) pages of your answer.*
4. Elaboration	4.1 Informative	*There is some research on Glucose Sensors already.*	*I especially liked the idea for thin-film nanosensor as a wearable device that can monitor human-body indices… seems that similar sensors already exist or are under development.*
	4.2 Suggestive	*…it would be creative, if the concept of an e-skin based glucose nanosensor could be elaborated.*	*Your thermometer will not measure accurately unless you can enter it inside the body.*

The distribution of the comment segments according to the feedback categories is presented in Figure 1. It shows that the *reinforcement* comments, categorized as low quality feedback, were less common among the in-class students (20.7%), compared to the MOOC students (44.5%). Regarding *clarification*, both groups indicated low percentages of comments, yet, with lower percentage among the in-class (0.9% and 2.8%). The *verification* comments, categorized as medium quality feedback, were more common among the in-class than the MOOC students (49.1% and 31.1%, respectively). Similarly, the *elaboration* comments, categorized as the highest quality feedback, were more common among the in-class than the MOOC students (29.3% and 21.6%, respectively).

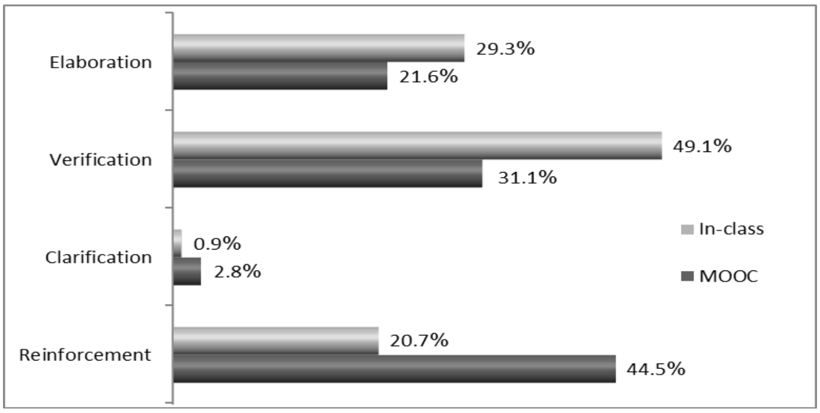

Figure 1. The distribution of the in-class and MOOC students' comment segments according to categories

The distribution of the comment segments according to sub-categories is presented in Figure 2. Interestingly, by examining the *reinforcement* subcategories, the in-class students have awarded their peers with very low percentage of negative comments (0.9%) compared to the MOOC students (3.2%). The latter provided relatively more negative comments, of which some were harsh and even insulting. Regarding *clarification*, the in-class students did not assert comments related to lack of scientific clarity, which suggests that the in-class assignments were scientifically well written and precise. *Verification* was more common among the in-class students with a relatively high percentage of validated work (38.8% validate, 10.3% invalidate) compared to the MOOC students who indicated a relatively high percentage of invalidated work (17% validate, 14.1% invalidate). Regarding *elaboration*, both groups indicated similar percentages of informative comments. However, suggestive comments, categorized as the highest quality feedback, were more common among the in-class students (21.6% and 14.5%, respectively). Analysis of the written feedback according to overall positive, neutral, or negative comments, indicated that students from both groups provided a similar percentage of positive comments (58.6% in-class and 58.3% MOOC). However, with regards to the negative comments, the MOOC students indicated higher percentages compared to their counterparts (17.3% and 11.2%, respectively).

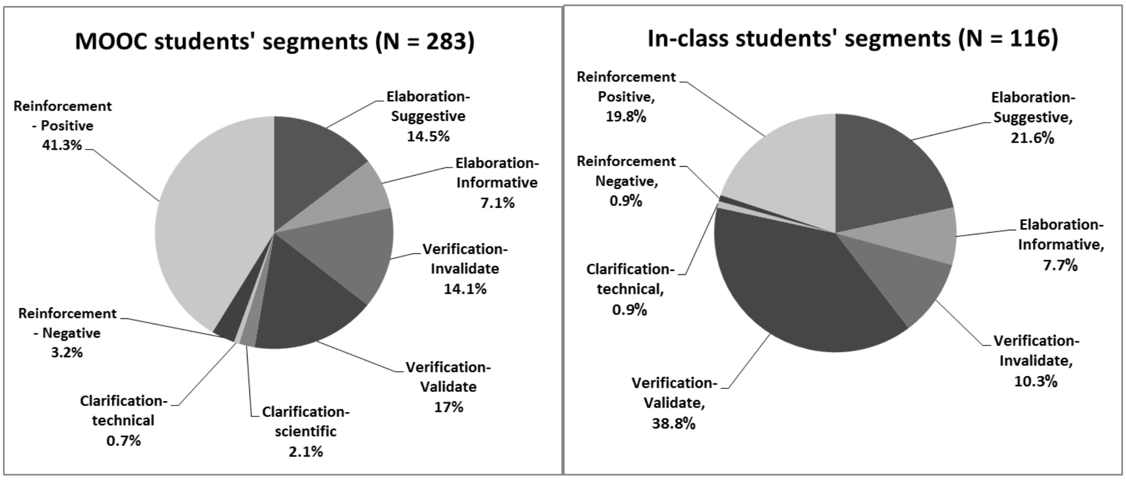

Figure 2. The distribution of the in-class and MOOC students' comment segments according to subcategories

Discussion

The current study examined the mechanism of online peer assessment (grading and feedback) while comparing between in-class and MOOC students. Findings indicated that the in-class students awarded lower grades to their peers compared to MOOC students and the TAs. The MOOC students awarded their peers similar mean grades to those given by the TAs, which might suggest that they were more "accurate"; however this was not the case. Findings indicated that only the grades of the in-class students were statistically positively correlated to those of the TAs. Hence, the in-class students were more stringent in grading their peer, but their grades were in good correlation with those given by the TAs.

In regard to our second research question, a higher percentage of the in-class students chose to provide written comments, compared to the MOOC students. Yet, out of those who did provide feedback, a higher number of segments per peer feedback was recorded among the MOOC students compared with the in-class students. A possible explanation for this could be the limited interactions between students and instructors in MOOCs; thus, the written comments might have filled this gap (Barak et al., 2016).

Evidence for higher order thinking was associated with the in-class students, compared to their MOOC counterparts. MOOC students provided their peers with more *reinforcement* and *clarification* comments, which were defined as lower order thinking. It seems that MOOC students were often confronted by lack of understanding of their peer work, which might be explained by the heterogeneity characteristics of MOOCs (Barak et al., 2016). Contrary to this, the in-class students provided their peers with more *verification* and *elaboration* comments, which were conceptualized as a more valuable feedback. This is persistent with the findings of Van der Pol and colleagues (2008), who reported that feedback in which learners suggest their peers with concrete ideas for revisions, holds the best potential for beneficial change. Overall, our findings indicated preference for the in-class group. However, due to the small number of participants, further research should be conducted for effective implementation of online peer assessment for both in-class and online learners.

References

Barak, M., & Rafaeli, S. (2004). Online question-posing and peer-assessment as means for Web-based knowledge sharing. *International Journal of Human-Computer Studies*, *61*(1), 84-103.

Barak, M. Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: examining aspects of language and social engagement. *Computers & Education, 94,* 49-60.

Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals*. *Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.

Çevik, Y. D. (2015). Assessor or assessee? Investigating the differential effects of online peer assessment roles in the development of students' problem-solving skills. *Computers in Human Behavior, 52*, 250-265.

Cheng, K. H., Liang, J. C., & Tsai, C. C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25*, 78–84.

Corder, G. W., & Foreman, D. I. (2014). *Nonparametric Statistics: A Step-by-Step Approach*, Wiley. ISBN 978-1118840313.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education & Training International, 32*(2), 175-187.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322.

Gielen, M., & De Wever, B. D. (2015). Structuring peer assessment: Comparing the impact of the degree of structure on peer feedback content. *Computer in Human Behavior, 52*, 315-325.

Hanrahan, J., & Isaacs, G. (2001). Assessing self- and peer assessment: The students' views. *Higher Education Research and Development, 20*(1), 53-70.

Kellogg, S., Booth, S., & Oliver, K. (2014). A social network perspective on peer supported learning in MOOCs for educators. *The International Review of Research in Open and Distributed Learning, 15*(5).

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction, 20*(4), 344-348.

Kulkarni, C., Wei, K. P., Le H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S. (2013). Peer and Self Assessment in Massive Online Classes. ACM Transactions on Computer-Human *Interaction (TOCHI), 20*(6), Article 33.

Kulkarni, C., & Socher, R., Bernstein, M., & Klemmer, S. (2014). *Scaling short-answer grading by combining peer assessment with algorithmic scoring*. Proceedings of ACM Learning at Scale. March, 2014.

Kulkarni, C., Bernstein, M., & Klemmer, S. (2015). *PeerStudio: Rapid peer feedback emphasizes iteration and improves performance*. Proceedings of the second ACM conference on Learning at Scale. March, 2015.

Luaces, O., Di'ez, J., Alonso-Betanzos, A., Trocoso, A., & Bahamonde, A. (2015). A factorization approach to evaluate open-response assignments in MOOCs using preference learning on peer assessments. *Knowledge-Based Systems, 85*, 322-328.

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology*, *36*(1), 91-115.

Sadler, P. M., & Good, E. (2006). The impact of self-and-peer-grading on student learning. *Educational , 11*(1), 1-31.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68(*3), 249-276.

Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers and Education, 51*, 1804-1817.

Yu, F. Y., & Wu, C. P. (2013). Predictive effects of online peer feedback types on performance quality. *Educational Technology & Society, 16*(1), 332-341.