

Research Papers

Prize as a Group Achievement: Incentivizing Accurate Peer Grading in MOOC

Nikan Sadehvandi, Toru Iiyoshi

(Center for the Promotion of Excellence in Higher Education, Kyoto University)

Peer assessment is the most widely applicable method of assessing students' open-ended assignments in MOOC. However, one of the most glaring problems with this type of assessment is lack of credibility of the grades. This study took a novel approach by using prize as an incentive to improve peer grading accuracy among MOOC learners, analyzing peer grading agreement, time spent on grading, and learners' perception of the grades. The results indicated that although the Virtual Prize Group (VPG) mechanism did not have any effect on improving peer grading agreement, it encouraged peer graders to spend more time on the grading activity. Also, most learners expressed their overall satisfaction with their grades and perceived them as fair. Finally, we provide directions for future research on increasing the quality of peer assessment through implementing similar attempts such as the one introduced in this paper.

Key words: MOOC, Peer assessment, Virtual Prize Group mechanism, Peer grading accuracy, Time spent on grading

1. Introduction

Peer assessment is vastly utilized in MOOCs as an attempt to introduce more depth into the evaluation process. It is also an alternative to compensate for a lack of instructor's grading and feedback on students' written works.

Conducting a good assessment requires spending an adequate amount of time on the part of learners. Sometimes, learners might conduct a cursory assessment of their peers' works if they do not feel the need to spend sufficient time reading them (Lu, Warren, Chaudhurim, & Rixner, 2015). If learners do not pay attention to the content of a submitted work, they might grade a submission poorly or generously and their grades do not reflect the true quality of that submission.

MOOC learners have different intents for registering in a course and not all of them plan to complete it. On the same note, many MOOCs strive to keep the attrition rate low and participation steady until the end of the course. Thus, penalizing poor grading activities might even lead to reverse results. After all, without ensuring peer graders that their effort in grading will be appreciated, it stands to reason that they do not take the assessment seriously. Thus, one way to value peer graders effort is to reward them. Prize incentives are deemed as rigorous extrinsic drivers

for students' success in a learning environment (Edmunds & Tancock, 2003; Xiong, Li, Kornhaber, Suen, Pursel, & Gions, 2015). Covington & Mueller (2001) state, "Extrinsic payoffs can either advance a love of learning—if they serve positive, task-oriented reasons—or interfere with caring if they are sought after for self-aggrandizing purposes" (p. 166).

Thus, prize may also lead to conflict among learners and interfere with creating an amiable environment where students can safely cooperate and share knowledge with each other. Thus, if peer graders are promised a prize at the end, they might no longer be completely impartial in their grading.

Thus, a better approach to utilize prizes in peer grading is to ensure that they contribute to motivating quality assessment on the part of learners. This study reports on such an attempt to improve the quality of peer grading through incentivizing MOOC learners to conduct a more accurate review of their peers' submissions, by taking a different approach to awarding a prize.

1.1. Why Peer Assessment

Peer assessment has been largely implemented as an alternative evaluation method in many higher education

settings worldwide (Wen & Tsai, 2006). The literature is replete with compelling arguments on the value of peer assessment as a learning tool that benefits both the reviewer and reviewee during a conventional peer assessment activity. The merits are manifold and often listed as promoting deep learning through reflection, raised autonomy, and fostering higher-order thinking skills (Topping, 1998). While saving teachers' time for post-assessment activities, peer assessment can boost students' meta-cognitive skills (Sadler & Good, 2006). If used as summative assessment method, peer assessment can also facilitate the development of new knowledge and skills (Cestone, Levine, & Lane, 2008).

1.2. Issues with MOOC Peer Assessment

However, the mechanics of peer assessment is far different in online courses where it is mainly used as a measurement tool. In MOOC, correcting and giving feedback on students' works clearly lead to an excessive workload for the instructor. Use of other assessment methods to replace instructor's feedback is also not uncommon in some MOOCs. For example, automated essay scoring utilizes several essay features to predict human grading on the same essay (Balfour, 2013). In the first glance, automated essay scoring might seem like a quick-fix, but it is by no means a one-size-fits-all solution when it comes to gauging learners' analytical reasoning and critical thinking. In fact, evaluating those essays that involves the use of such skills can be quite subjective and topic-dependent. Moreover, such works also call for assessor's higher level of thinking to assess the given work on a deeper level than merely attending to the syntactic structure or relying on word count. Thus, many a MOOC rely on peer assessment as a solution to measure students' deeper understanding of the subject as presented in their essays.

However, even when MOOC peer assessment is carefully designed and carried out, certain problems might arise. These problems might stem from a range of possible reasons such as lack of pertinent knowledge and expertise, lack of willingness to put sufficient effort into the review process, or even maliciousness. They may be eventually manifested in the form of colluded or rogue assessment on the part of learners (Knight & Steinbach, 2011).

Rogue assessment is the result of peers' indolence, partiality, or other similar factors which prompt learners

to conduct an inadequate assessment on their peer's work. These factors are more aggravated in online peer assessment since learners barely identify themselves as part of a learning community (Lu & Bol, 2007). As in traditional classes, in a typical MOOC peer grading activity, the identity of learners is not revealed to the corresponding peer, and vice versa. It is argued that anonymity alleviates learners' unjust grading behavior (MacLeod, 1999). Others suggest that students usually tend to give generous scores on their friends' essays when peer graders are identified (Vickerman, 2009). In fact, friendship might result in dishonest grading, as students might feel compelled to provide inaccurate, inflated, or deflated grades (Burdett, 2007; Llado, Soley, Sansbello, Pujolras, Planella, Roura-Pascual, Martinez, & Moreno, 2013). In traditional classes with the instructor being frequently present, anonymous grading activities can lessen partial grading behavior triggered by comradeship (Darling-Hammond, Ancess, & Faulk, 1995).

However, anonymity works conversely in MOOC because students do not feel the obligation to be sedulous and accountable with their grading. In the absence of instructor or a screening authority coupled with less perceived degree of social presence, students might even find little worth in investing time and effort on peer grading. Also, students' grading behavior does not affect their social face since they do not have a strong sense of community affiliation (Lu & Bol, 2007).

MOOC research is beginning to realize the importance of incentivizing peer graders in order to conduct a better assessment (Piech, Huang, Chen, Do, Ng, & Koller, 2013). Similarly, a few studies have tested the effects of hands-on models such as "organic peer assessment" web-based tool, "grading the grader" regime (Lu et al., 2015) and "identified peer reviewing" framework (Gamage, Whiting, Rajapakshe, Tilakaratne, Perera, & Fernando, 2017) to motivate peer graders and enhance the quality of peer assessment. According to Lu et al. (2015), one obvious reason for students' low-quality grading is lack of sufficient effort spent on the assessment activity. Lu and colleagues maintain that to give motivation rather than punishment to peer assessors for doing a better job on assessment can have a considerable effect on the quality of peer assessment results. Designing and implementing a "grading the grader" regime in a MOOC peer assessment activity on Coursera, Lu and his colleagues concluded that informing learner

graders that the submissions they graded would be regarded by others can inspire more deliberate efforts towards assessment quality. In a newly published pilot study on testing the effects of an Identified Peer Reviewing (IPR) framework, Gamage et al. (2017) concluded that revealing the identity of peer assessors as opposed to having them conduct a blind review is more effective.

However, making peers' identity known to other learners is not an option under the students' privacy policy adopted by most MOOC platforms. Given the circumstances, perhaps a better approach is one that increases a sense of group belonging, while keeping the anonymity feature functioning.

1.3. Virtual Prize Group Mechanism (VPG)

This study investigates the effects of an incentivizing mechanism, dubbed virtual prize group (VPG), on improving the quality of peer assessment in MOOC. Through taking a different approach to utilizing the otherwise equivocal anonymity feature of MOOC combined with *best performance award* offered to a group of peer graders, the mechanism has two purposes. First, it aims to improve the accuracy of peer awarded grades on a homework assignment. Secondly, it seeks to encourage students to spend more time on the peer grading activity.

The core attribute of the VPG mechanism was to allocate a prize on a group-based format as opposed to awarding it to individual peer graders while keeping the identity of the groups anonymous. Learners who submitted their work were informed prior to the advent of peer assessment that they each belonged to a randomly predefined group and only one of these groups would receive the prize as a group-based accomplishment. However, the identity of the group members was neither divulged to the rest of the group nor other learners in the other rival groups. The assignment of multiple learners to grade a single submission is conducted by MOOC default peer review system with the anonymous feature being enabled. Even without making the identity of learners in the virtual prize groups anonymous, learners had little chance of knowing who would grade a given submission.

However, to further ensure that learners did not confirm if a given essay was submitted by one of their own virtual prize group members through the discussion forum or other possible channels, the researcher kept the members'

identity confidential until the end of the peer assessment activity.

As shown by the example in Figures 1, learner A is required to grade both learner B and learner C while learner A has no knowledge of whether these two learners are sharing the same virtual prize group with him. Thus, learner A would do his best effort to neither underscore nor over-score these learners. Supposedly, if learner A is not aware of the VPG mechanism and gives a deflated grade to learner B, it might affect the entire group's performance to which learner A belongs as well. Similarly, if learner A inflates learner C's score, who is from a different group, the same consequence may occur. Therefore, not having any knowledge of who the other members of his virtual prize group are, learner A tends to be as attentive as possible in his grading because he has his own group's interest at stake.

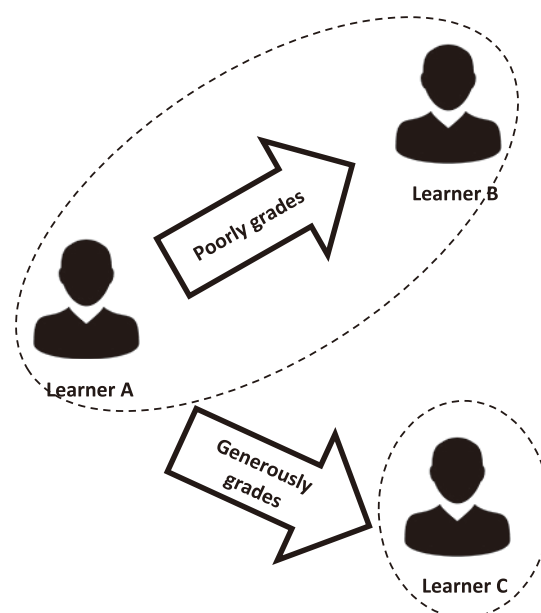


Figure 1. Peer grading in virtual prize group mechanism*

*Virtual prize groups are presented by dotted circles.

2. Methodology

2.1. Study context

The study was conducted in one of the offerings of KyotoUx: "Culture of Services: New perspective on customer relation" (002x) MOOC. The course was released over eight weeks in 2016. The virtual prize group mechanism was embedded for peer assessment on homework 5-2 in week 6. The homework was an extension of homework 5, which was the last weekly homework of the course and counted for 15% of the final grade.

The homework required students to watch a video clip and capture different aspects of interaction between a customer and employee in a fast food restaurant. Students needed to describe these aspects based on the following prompts and submit their answers for grading by the designated deadline.

A. Describe the customer's gesture when conveying the order.

B. Describe where she looks when conveying the order.

You can discuss what she does not look at.

C. Describe her talk of conveying the order.

Each prompt contained 5% of the final grade. Each learner was required to conduct a minimum of two peer reviews, using a pre-designed grading rubric along with the provision of ideal responses on each homework questions.

2.2. Peer grading rubric

The peer grading rubric was generated by the instructor of KyotoUx (002x). The rubric was supplemented with the instructor's ideal responses to each of the three prompts described in the previous section (Table 1).

Table 1. Instructor's ideal response to each prompt

	Homework prompts	Instructor's sample responses	
A.	Describe the customer's gesture ...	Important points	She points at the area of the menu where ...
		Additional points	She first points at the big image of the burger ...
B.	Describe where she looks ...	Important points	She looks only at the menu, and not at the employee ...
		Additional points	She moves her face closer when she say ...
C.	Describe her talk ...	Important points	She was saying the name of the burger slowly ...
		Additional points	Before saying the name of the burger, she prefaced it by ...

The rubric also contained two questions which worked in concert with the instructor's ideal responses. These two questions asked whether the submission at hand captures any of the important or additional points in the instructor's ideal responses:

- 1. Does your peer's response to this homework capture the important points in the instructor's ideal response?*
- 2. Does your peer's response to this homework capture the additional points in the instructor's ideal response?*

In order to assess to what extend a submission captures the important and additional points in the instructor's ideal response, the following scales were created:

- *Good* (important points, 3 pts; additional points, 2 pts): the response completely captures all the points in the instructor's sample response.
- *Satisfactory* (important points, 2 pts; additional points, 1 pts): the response partially captures the points in the instructor's sample response
- *Needs improvement* (important points, 1 pts; additional points, 0 pts): the response does not capture any of the points at all.

Moreover, there was an instruction video and a training step on how to optimally use the peer grading rubric. This instruction video was accessible up to the end of the course. Thus, students could refer to it anytime during the peer assessment activity if they chose to.

2.3. Allocation of prize

The prize was designated for the best performance on homework 5-2. Ultimately, only one group whose sum of members' grades on this homework was higher than the other groups was selected as winner and awarded the prize at the end of the peer assessment process. Simultaneously with the launch of homework 5-2 on the courseware, students were informed about the virtual prize group mechanism through the instruction video for homework 5-2, course update, newsletters, and the discussion forum specified thread.

As for prize, *Recognition of Best Performance on Homework 5-2 in Culture of Services: New perspective on customer relations* was developed by the KyotoUx design team and signed and granted by the course instructor. Ten electronic copies of this prize were sent to the members of the winner group via email.

The virtual grouping for allocation of the prize entailed IDs of those learners who submitted homework 5-2. Each group included ten learners who were selected through stratified sampling. The sampling was conducted manually based on learners' course performance report from the beginning of the course up to the start of week 6 as a proxy to define different performance strata. This report included learners' grades on weekly problem sets, completion checklists and homework assignments, which counted for 30%, 5% and 29% of the final grade,

respectively. The reason for using stratified sampling rather than normal random sampling was to give all learners including low-performers as well as high-performers an equal chance to win the prize at the end. The sample was constructed from a total of four strata: high-performers ($n = 56$), intermediate-performers ($n = 61$), low-performers ($n = 29$), and extremely low-performers ($n = 34$).

2.4. Course participants

The main goal of this study was to determine whether there was some degree of evidence that students, who were informed about the VPG mechanism, did better on grading their peers. If so, then it would be possible to argue that such a mechanism might be effective to encourage more accurate peer awarded grades.

180 learners took part in the peer grading activity for homework 5-2. Overall, they conducted 390 peer grading activities, mostly contributing two grades per submission. These learners had to submit their own works first in order to receive submissions to grade. Of this number, 95 peer graders were corroborated to have been cognizant of the VPG mechanism while 77 peer graders were corroborated to have been incognizant.

These two groups will be referred to as “informed” and “uninformed”, respectively. The remaining eight peer graders could not be assigned to any of the two groups. Grouping of peer graders as either “informed” or “uninformed” was carried out utilizing students’ click data and access date for the video instruction on homework 5-2. This data was further supplemented by students’ responses on whether they knew about the possibility of receiving a prize before the start of peer grading activity.

As was mentioned earlier, although the minimum requirement for learners was to grade two submissions, there were also cases of grading more than two submissions by some peer graders. However, since these cases were few in number, the analysis only contained 344 grades, awarded by 172 pairs of peer graders, each pair grading a single submission. These consisted of 68 pairs of grades by the informed condition, 36 pair of grades by the uninformed condition, and 68 pairs of grades awarded by one informed and one uninformed peer grader. This conditioned was referred to as “mixed”. The analysis was conducted comparing these three conditions. Both peer graders were randomly distributed in the three conditions. Therefore, as

conceptualized in Figure 2, for example, an informed peer grader from the informed condition might have graded another submission in the mixed condition as well. Also, an uninformed peer grader from the uninformed condition might have graded another submission in the mixed condition as well.

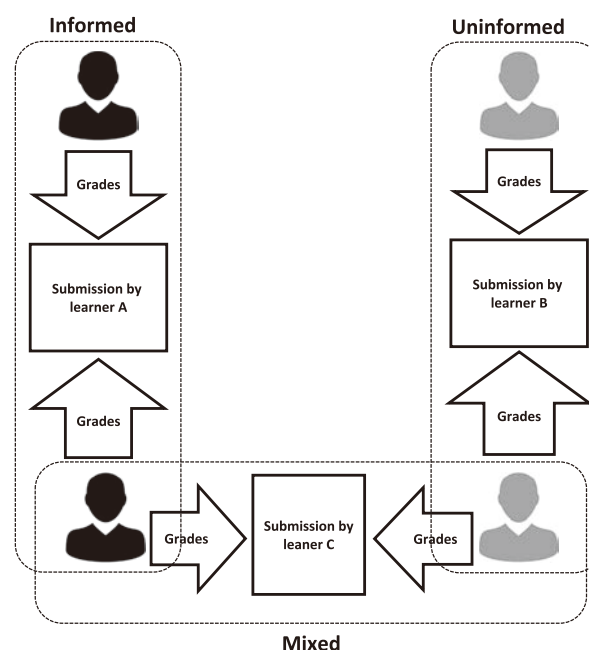


Figure 2. The three conditions

2.5. Data sources

Several survey questions were created and imbedded onto the course ending survey. The first survey question asked students about the average amount of time spent on grading the two submissions using a drop-down menu. Also, students were asked whether they were cognizant of the possibility of receiving a prize for best performance on homework 5-2 when they did peer grading for this homework, choosing either “Yes, I did.” or “No, I did not.” from a drop-down menu. Those who had selected the first response were additionally asked to specify which feature of the courseware (i.e. homework 5-2 instruction video, discussion forum, course update, or Newsletter) informed them about the prize.

In order to grasp learners’ perception about the fairness of the grades that they received from their peers and their level of satisfaction with those grades, the following survey items were created and imbedded in the course ending survey:

1. I am satisfied with the grades that I received from my peers for my homework 5-2.

2. *I believe that all peer assessors graded my homework 5-2 fair and reasonably.*

77 learners responded to these two items, selecting among seven options ranging from “1-strongly disagree” to “7-strongly agree” to provide their responses on these items.

In order to find out which condition (i.e. informed, uninformed, or mixed) the respondents received their two grades from, all the respondents’ IDs were matched with their corresponding peer graders’ IDs. Based on this, the respondents divided to 30 learners graded by informed peer graders, 17 learners graded by uninformed peer graders, and 30 learners graded by one informed and one uninformed (i.e. mixed) peer graders.

2.6. Data analysis

This study implemented three proxies to operationally define accuracy. First, the researcher assumed that students’ grades on the same submission tend to be in more agreement with each other if they are similarly motivated. In order to establish if the VPG mechanism was effective to raise the level of agreement between peer graders, Intraclass Correlation Coefficient (ICC): the one-way random effects (1, 2) model (absolute agreement) was conducted between the two sets of grades given by the three conditions (Shrout & Fleiss, 1979). This model corresponds well to the characteristics of peer assessment in this study, that is, the peer graders were randomly selected from a larger pool and each submission was graded by a set of different peer graders. According to Koo & Li (2016), the following formula is used in conjugation with this model:

$$\frac{MS_S - MS_{PG \times S}}{MS_S}$$

In this setting, MS_S is means square for the submissions, and $MS_{PG \times S}$ is the means square for the interaction between peer graders and submissions. Previously, Luo, Robinson, & Park (2015) successfully used the same model of ICC to calculate the interrater reliability of peer’s submitted grades on Coursera platform.

The second proxy used to establish accuracy in this study was the amount of time students spent on grading submissions. As was mentioned earlier, spending short duration of time on reviewing a submission leads to a poor

grading activity and encouraging the allocation of more time could raise peer assessment quality (Piech et al., 2013). Similarly, the researcher assumed that if the informed peer graders are sufficiently motivated to be more attentive to their grading, they tend to spend longer duration of time grading a given submission compared to their uninformed peers. The amount of time spent on grading was elicited from learners through a survey question. Chi-square was used to find any significant differences between the average amount of time reported by the informed and uninformed groups on grading the submissions.

Finally, it was assumed that those learners who are graded by the informed condition tend to be more satisfied with their grades and perceive them to be fair compared to those who are graded by either uninformed or mixed conditions. To this end, frequency analysis and Chi-square were used to test any significant differences between responses from learners graded by either of the three conditions with respect to their satisfaction and perception of fairness.

3. Results

3.1. Comparing grades across the three conditions

Table 2 shows the descriptive statistics for the three conditions. Among the three conditions, the informed condition was shown to have the highest mean score whereas the uninformed condition had the lowest mean score.

Table 2. Descriptive statistics

Groups	<i>N</i> *	<i>M</i>	<i>SD</i>
Uninformed	36	9.80	3.67
Informed	68	11.11	3.06
Mixed	68	10.45	3.95

*Number of grading contributions by the two peer graders

Furthermore, the average measure ICC (.316) for the informed condition is the lowest among the three (Table 3). This ICC is of a poor strength (Koo & Li, 2016). The average measure ICC (.461) for the uninformed condition is also of a poor strength although, it is slightly higher than that of the informed. Also, the highest average measure ICC is for Mixed condition (.595), which is of a moderate strength.

The results of the analysis of ICC also indicated that while the agreement between the two peer graders in the

informed condition is not statistically significant, they are significant between the two peer graders in the uninformed and mixed conditions. Based on these results, it can be concluded that the VPG mechanism did not have any effects on bringing informed students' graders in higher agreement as compared to other conditions.

Table 3. Intraclass correlation between the two peer graders in each condition

Conditions	Informed (<i>n</i> = 68)	Uninformed (<i>n</i> = 36)	Mixed (<i>n</i> = 68)
Single measures	.181	.302*	.410*
Average measures	.316	.461*	.595*

* $p < 0.05$

3.2. Findings from the survey

(1) Time spent on grading

Figure 3 shows the frequency analysis of average amount of time spent on grading as reported by the learners.

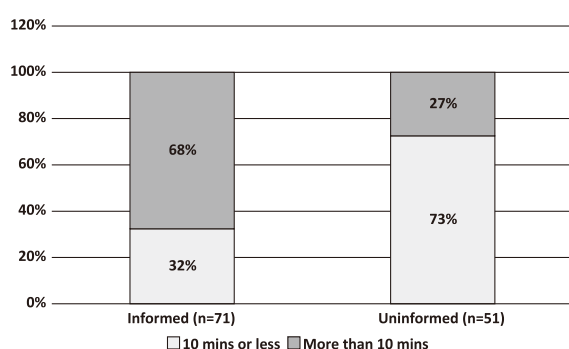


Figure 3. Time spent by informed and uninformed groups on peer grading

As one can observe, 48 peer graders (68%) in the informed group reported to have spent more than 10 minutes on grading the submissions while 37 peer graders (more than 70%) in the uninformed group reported to have spent less than 10 minutes on grading.

The findings from a chi-square test also indicated that there was a significant difference between informed and uninformed students with respect to the average amount of time that they spent on grading their peers ($\chi^2(1, n = 122) = 19.96, p = .000, \phi = .41$), with the informed group reporting to have spent more time than their uninformed peers.

(2) Learners' perception

Table 4 shows the frequency analysis as well as mean score for learners graded by each condition. More than 80%

of the respondents graded by each condition agreed that their grades were fair and they were satisfied with them.

Table 4. Proportion of dis-/agreement on survey Items 1 & 2

Survey items	Graded by	N	Disagree	Agree	M
1.	Informed	30	11%	89%	5.8
	Uninformed	17	13%	87%	6.2
	Mixed	30	9%	91%	5.9
2.	Informed	30	15%	85%	5.2
	Uninformed	17	20%	80%	5.5
	Mixed	30	11%	89%	5.3

Furthermore, the findings from a chi-square test showed that there was no significant difference among the three categories with respect to their level of satisfaction with the grades ($\chi^2(1, n = 97) = 9.66, p = .64, \phi = .31$) and their perception of fairness ($\chi^2(1, n = 97) = 12.43, p = .41, \phi = .33$).

4. Discussion

The study detailed the effort at designing and utilizing a mechanism to increase the accuracy of peer grading in one of the offerings of KyotoUx. Accuracy is a rather subjective trait and sometimes difficult, if not impossible, to establish in a course of a single intervention. In this study, peer grading accuracy was defined through looking into several possible changes that might occur in peer awarded grades as a result of implementing the VPG mechanism. As was previously reported by the findings, the VPG mechanism did not have any significant effect on increasing the level of agreement between the two peer graders in the informed condition. Also, the descriptive statistics demonstrated that the informed condition ($M = 11.11$) tended to give slightly more generous grades on the submissions than the uninformed condition ($M = 9.80$). Findings from the ending survey also indicated that the peer awarded grades were well received by the learners, that is, they were mostly satisfied with the grades and perceived them as fair.

However, given the results of the Chi-square test which showed that the informed group spent more time on peer grading activity, it could be concluded that the mechanism might have impacted the motivation of this group to spend longer on grading the submissions than the uninformed group.

The argument supporting the need for motivating learners with peer grading has long been made clear by previous research, both in the context of traditional classroom as well as online learning (Higgins, Hartley, & Skelton, 2002; Piech et al., 2013). In fact, this study provided empirical evidence that incentivizing peer graders could encourage learners to be more dedicated to their grading. The fact that the informed peers spent significantly greater amount of time on grading than the uninformed peer graders might also indicate more deliberate effort towards generating accurate grading results.

However, it might also signal the existence of other potential problems. For instance, peer graders might have experienced more difficulties with grading the submission at hand. According to Freeman & Parks (2010), “students not only have trouble answering high-level application, analysis, synthesis, and evaluation questions, they also have trouble grading them” (p. 487). In this sense, while the amount of time spent on grading might indicate that the informed group invested more time on the grading, they might not have been able to give more accurate grades on the submissions than their uninformed peers, as the findings from the analysis of ICC suggested. As stated earlier, in order for peer graders to upgrade their peer grading skills, there was some supplementary training materials (i.e. an instruction video and a training step) imbedded within the courseware. Every student, informed as well as uninformed, could have referred back to them in order to grasp a better understanding of the grading rubric. However, questions of whether the mechanism prompted students to make use of the training material in order to improve their assessment skills or even how spending longer on grading might have fed into students’ learning was not focused in this study.

5. Future Research

One inherent challenge with implementing the mechanism was the way it was introduced within the courseware. As previously mentioned, some of the available features of the courseware were used to notify learners about the competition-like add-on (i.e. the VPG mechanism) into the peer assessment module. However, whether students paid sufficient attention to the information provided through these features leaves room for speculation. Also, it is not clear whether the informed learners shared this information with their uninformed friends later on throughout the

peer assessment activity. Therefore, a better method for implementing such a mechanism, which eventually needs some degree of cooperation on the part of learners, is to notify them weeks prior to its implementation. This could be done together with a promise of worthwhile prize that offers strong impetus to students for performing an attentive peer grading activity.

Also, it will be helpful to investigate the effectiveness of the mechanism based on the extent to which students might value the prize internally and endeavor towards accuracy with the hope of receiving a prize at the end.

6. Conclusion

Learners might often doubt the quality of peer assessment and their peers’ knowledge of what constitute a sound assessment practice (Kaufman & Schunn, 2010). Awards incentives can help increase learners’ overall satisfaction with peer assessment outcomes as they can assure students that their effort is not neglected and measures have been taken to fine-tune the peer assessment activity in order to achieve more reliable assessment results.

Attempts to monitor students’ attentiveness to peer assessment—such as the one described in this paper—could also benefit learning. Having students spend sufficient amount of time on a given submission might help them reach a deeper level of reflection, which might also lead to their better understanding of the subject and reshaping their knowledge. Finally, if implemented in ideal form, VPG mechanism might contribute to a greater sense of accountability in assessment, which is, otherwise, low among MOOC peer graders.

References

- Balfour, S. P. (2013). Assessing writing in MOOCs: Automated essay scoring and calibrated peer reviewTM. *Research & Practice in Assessment*, 8.
- Burdett, J. (2007). Degrees of separation –balancing intervention and independence in group work assignments. *The Australian Educational Researcher*, 34(1), 55–71.
- Cestone, C. M., Levine, R. E., & Lane, D. R. (2008). Peer-assessment and evaluation in team-based learning. *New Directions for Teaching and Learning*, 11(6), 69–78.
- Convington, M. V., & Mueller, K. J. (2001). Intrinsic versus extrinsic motivation: An approach/avoidance

- reformulation. *Educational Psychology Review*, 13(2), 157–176.
- Darling-Hammond, L., Ancess, J., & Faulk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York: Teacher College Press.
- Edmunds, K. M., & Tancock, S. M. (2003). Incentives: The effects on the reading motivation of fourth-grade students. *Reading Research and Instruction*, 42(2), 17–38.
- Freeman, S., & Parks, J. W. (2010). How accurate is peer grading? *CBE—Life Science Education*, 9(4), 482–488.
- Gamage, D., Whiting, M. E., Rajapakshe, T., Thilakarathne, H., Perera, I., & Fernando, S. (2017). Improving assessment on MOOCs through peer identification and aligned incentives. *Proceedings of the Fourth ACM Conference on Learning @ Scale*, 315–318.
- Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53–64.
- Kaufman, J. H., & Schuun, C. D. (2010). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science*, 39(3), 387–406.
- Knight, L. V., & Steinbach T. A. (2011). Adapting peer review to an online course: An exploratory case study. *Journal of Information Technology Education*, 1, 81–100.
- Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Llado, A. P., Soley, L. F., Sansbello, R. M. F., Pujolras, G. A., Planella, J. P., Roura-Pascual, N., Martinez, J. J. S., & Moreno, L. M. (2013). Student perceptions of peer assessment: an interdisciplinary study. *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2013.860077
- Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning*, 6(2), 100–115.
- Lu, Y., Warren, J., Jermaine, C., Chaudhuri, S. & Rixner, S. (2015). Grading the graders: Motivating peer graders in a MOOC. *Proceedings of the 24th International Conference on the WorldWide Web*, 680–690.
- Luo, H., Robinson, A. C., & Park, J. Y. (2015). Peer grading in a MOOC: Reliability, validity, and perceived effects. *Online Learning Journal*, 18, 1–14.
- MacLeod, L. (1999). Computer aided peer review of writing. *Business Communication Quarterly*, 62(3), 87–95.
- Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. Retrieved from: <https://web.stanford.edu/~cpiech/bio/papers/tuningPeerGrading.pdf> (31 August, 2018)
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1–31.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428.
- Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 63(3), 249–276.
- Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education*, 34(2), 221–230.
- Wen, M. L., Tsai, C. C. (2006). University students' perceptions of and attitude toward (online) peer assessment. *Higher Education*, 51, 27–44.
- Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B. & Gions, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: a structural equation modeling approach. *Global Education Review*, 2(3), 23–33.