Exploring the Role of Online Peer-Assessment as a Tool of Early Intervention

Michael Mogessie Ashenafi[(✉)], Marco Ronchetti,
and Giuseppe Riccardi

Department of Information Engineering and Computer Science,
University of Trento, Trento, Italy
{michael.mogessie, marco.ronchetti,
giuseppe.riccardi}@unitn.it

Abstract. Peer-assessment in education has a long history. Although the adoption of technological tools is not a recent phenomenon, many peer-assessment studies are conducted in manual environments. Automating peer-assessment tasks improves the efficiency of the practice and provides opportunities for taking advantage of large amounts of student-generated data, which will readily be available in electronic format. Data from three undergraduate-level courses, which utilised an electronic peer-assessment tool were explored in this study in order to investigate the relationship between participation in online peer-assessment tasks and successful course completion. It was found that students with little or no participation in optional peer-assessment activities had very low course completion rates as opposed to those with high participation. In light of this finding, it is argued that electronic peer-assessment can serve as a tool of early intervention. Further advantages of automated peer-assessment are discussed and foreseen extensions of this work are outlined.

Keywords: Electronic peer-assessment · Early intervention · At-risk students

1 Introduction

Peer-assessment in education is an assessment method in which students assess the performance of their peers. Topping [1] defines peer-assessment more formally as "…an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status."

A wide variety of peer-assessment settings exist in which the nature of the work being assessed varies with the discipline and course. Essays, answers to open questions, and oral presentations are examples of work that is assessed in peer-assessment classes.

Reliability, validity, and practicality of peer-assessment as well as its impact on students' learning have been studied for decades. Nonetheless, there is no strong consensus among practitioners on whether peer-assessment is guaranteed to deliver its

desired effects. Most recent research seeking to establish whether peer-assessment can be used as alternative assessment method has come in the form of one-off experiments conducted in small classes.[1]

The wide range of scenarios in which peer-assessment is implemented has also made it particularly difficult to reach solid conclusions about its effectiveness as the number and impact of variables being studied varies from one scenario to another. Some variables are however common to all peer-assessment settings. Examples include the number of students involved per assessment task and the total number of participants. Recent studies in peer-assessment have focused on a number of themes including peer-feedback, design strategies, students' perceptions, social and psychological factors, and validity and reliability of the practice.

Large-scale introduction of peer-assessment in educational institutions is very rare for a number of reasons. Given the issues of reliability and validity of student ratings, in particular, integrating peer-assessment into a course's curriculum implies risks both on parts of the institution and students, which stakeholders may not be willing to assume.

Moreover, the manual nature of most peer-assessment practices prevents researchers from carrying out large-scale experiments that would help investigate the impact of variables on the effectiveness of the practice.

Automation of peer-assessment tasks could greatly improve efficiency and enable conducting large-scale, iterative experiments. In addition to the foreseen improvement in efficiency, the move towards automated peer-assessment could reduce, if not eliminate, issues of non-confidentiality, the potential for academic dishonesty, and increase in the workload of instructors.

Automated peer-assessment implies more than simple automation of task assignment and submission. Depending on the nature of the work being assessed, many other features of peer-assessment can also be automated, at least to a certain degree.

Other opportunities that arise from the automation of peer-assessment tasks include mathematical modelling of students and construction of student profiles, ubiquitous peer-assessment practices that go beyond the confines of the traditional classroom, and creation of platforms that allow easy replication and extension of previous studies.

Another prospect worth examining is the potential of online peer-assessment systems to serve as tools of monitoring and supervision of students. Used this way, an online peer-assessment system may provide timely information about students who may be at risk of falling behind or even failing a course.

Two previous studies demonstrated how models of student progress and early intervention could be built on top of an online peer-assessment platform to monitor students who often participate in peer-assessment activities [4, 5]. This study is a follow-up, which sought to determine whether analysis of the digital traces of students who had little participation in peer-assessment activities could lead to a firm conclusion about their success rate of completing their courses.

---

[1] See [2] for a meta-analytic review of peer-assessment studies conducted in the previous century and [3] for a comprehensive review of those peer-assessment studies conducted since 2000.

The findings of three courses involving over 600 students, in which the online peer-assessment system was applied, showed that students who had a participation rate of less than 33% in online peer-assessment activities were significantly more at risk of not completing the courses as those who had over 33% participation rates.

Section 2 provides a brief review of the state-of-the-art in automated peer-assessment, with focus a number of popular tools for peer-assessment and recent topics addressed by the automated peer-assessment community. Section 3 introduces the reader to the peer-assessment system that was used to conducted this study. Section 4 presents analyses and findings across three computer science courses that were taught between mid-2013 and mid-2014. Section 5 provides a discussion of the role automated peer-assessment may play in identifying at-risk students and concludes with a number of final remarks.

2 Previous Work in Automated Peer-Assessment

Peer-assessment has been in use at all levels of education for over half a century. There is a significant amount of literature that documents this use across several course subjects. Writing courses, especially those that focus on improving students written English, constitute the largest share [2, 3]. Adoption of information technology infrastructures geared towards education by many institutions has led to the introduction of electronic peer-assessment systems with a varying designs and levels of sophistication. Educators in computer science have especially benefited from this transformation by introducing such tools in their courses.

There is an extensive number of tools that support automated peer-assessment. For reasons of brevity, four tools that are similar to the peer-assessment system used in this study are presented here.[2]

2.1 PRAISE (Peer Review Assignments Increase Student Experience)

de Raadt et al. [7] presented a generic peer-assessment tool that was used in the fields of computer science, accounting and nursing. The instructor could specify criteria before distributing assignments, which students would use to rate their peers' assignments. The system could compare reviews and also suggest a mark. Disagreements among reviewers would lead the system to submit the solution to the instructor for moderation. The instructor would then decide the final mark.

2.2 PeerWise

Denny et al. [8] presented PeerWise, a peer-assessment tool, which students used to create multiple-choice questions and answer those created by their peers.

---

[2] See [6] for a comprehensive review of tools that support peer-assessment.

When answering a question, students would also be required to rate the quality of the question. They could also comment on the question, in which case the author of the question could reply to the comment.

2.3 PeerScholar

Paré and Joordens [9] presented another peer-assessment tool, which was initially designed with the aim of improving writing and critical thinking skills of psychology students. First, students would submit essays. Next, they would be asked to anonymously assess the works of their peers, assign marks between 1 and 10, and comment on their assessments. An additional feature of PeerScholar is that students could also rate the reviews the received.

2.4 Peer Instruction

Peer Instruction is not a software artifact but a practice that is applied in a classroom setting where instructors provide students with in-lecture multiple choice questions. Students would then vote for the correct answer using Electronic Voting Systems (EVS), also known as clickers [10, 11]. A study by Kennedy and Cutts [12] showed that communication systems such as EVS may serve as indicators of expected student performance. This is an example of assessment that aids in early discovery of challenges students may face in grasping concepts so that appropriate supervision is provided in a timely manner.

Despite its widespread use, peer-assessment had not gained enough of the spotlight to warrant the creation of conferences or workshops dedicated solely to the topic. Now, there are at least two annual workshops that aim to bring together scholars from many disciplines in order to foster the advancement of the practice. The Computer Supported Peer Review in Education (CSPRED) and Peer Review, Peer Assessment, and Self Assessment in Education (PRASAE) workshops address several issues in electronic peer-assessment such as improving the impact of reviews [13, 14], training reviewers to improve the peer-review process [15, 16] and also aim to provide the current state-of-the-art in electronic peer-review and peer-assessment [17, 18].

3   The Online Peer-Assessment System

A web-based peer-assessment platform was developed and used in three computer science courses between 2012 and 2016. Participation in peer-assessment tasks was optional. However, all students who completed at least a third of the tasks were awarded a bonus worth 3.3% of the final mark. An additional 3.3% bonus was awarded to the top-third students, based on the number of peer-awarded points. For all three courses, it was observed that active participation in peer-assessment activities waned towards the end of the course. Regardless. A total of 83% of students for the three courses completed at least a third of the tasks.

In the latest version of the platform, students completed a weekly peer-assessment cycle composed of three tasks. First, students submitted questions relating to a list of previously discussed topics provided by the teacher. Then, the teacher examined the questions and selected a subset of questions, which were randomly assigned to students in the second task. The assignment procedure ensured that each question was assigned to at least four students. Machine learning techniques were applied to group similar questions in an effort to facilitate the question selection process.

After answers were submitted, question-answer sets were randomly distributed to students to rate each answer. In order to encourage careful assessment of each answer, students were provided with a certain amount of points, referred to as coins, to distribute over the answers.

After the completion of each cycle, questions and answers, together with coins earned, were made available to all students. Students could also monitor their progress by accessing visual and statistical information available in their profile page. In previous studies, how to predict expected performance in final exams was explored using student activity data from the peer-assessment system [4, 5]. The prediction models in those studies, however, considered only those students who had completed over a third of all peer-assessment tasks. The main reason behind this was that the performance of the linear regression models was significantly reduced with the introduction of data of students with little or no participation at all. Because the number of students who did not participate enough in online peer-assessment tasks was considerably low, the attempt to build prediction models only for those students did not produce encouraging results. Therefore, the analyses presented here used less sophisticated statistics to perform comparisons between the two student groups.

4   Analyses and Results

Three undergraduate-level computer science courses, labelled **IG1, LP,** and **PR2**, utilised the peer-assessment system between the early 2013 and mid 2016. The courses were administered at the University of Trento in Italy.

The Italian grading system uses a scale that ranges between 0 and 30, with 30L or 30 Excellent the highest possible mark. In order to pass a course, students have to obtain at least a score of 18. For the purpose of this analysis, the range of scores was categorised into four groups and labels were assigned to each group. Table 1 presents this partitioning of scores. The Italian higher education system permits students to sit for the same exam at most five times within an academic year. Students therefore have an opportunity to improve their grades by making several attempts. The analysis considers the data of those students who both subscribed for peer-assessment tasks and attempted an exam at least once.

For each course, students with less than 33% participation (who completed less than a third of peer-assessment tasks), referred to as *low-participation groups*, and those with over 33% participation, referred to as *high-participation groups*, were assigned to performance groups according to the final scores they obtained for the courses. Table 2 shows the categorisation of students in the low-participation group

Table 1. Mapping of scores to performance groups

Score	Verdict
Below 18	Insufficient
Between 18 and 22	Low performer
Between 23 and 26	Medium performer
27 or above	High performer

into the performance groups and Table 3 does the same for those in the high-participation group. The observation that a large majority of low-participation students did not manage to obtain passing marks was consistent across all three courses, albeit with varying degrees. In the case of PR2, low-participation students were more than twice as likely to score below the passing mark as those with high participation. LP and IG1 low-participation groups were 1.66 and 1.87 times as likely as their high-participation counterparts to score below the passing mark, respectively.

Table 2. Distribution of scores for students in the low-participation group

Course	Number of students	<18	[18.23)	[23, 26]	>=27
IG1	35	25	4	3	3
PR2	42	18	7	11	6
LP	30	22	1	5	2

Table 3. Distribution of scores for students in the high-participation group

Course	Number of students	<18	[18.23)	[23, 26]	>=27
IG1	182	69	46	36	31
PR2	141	23	33	49	36
LP	189	84	40	37	28

Further analysis of the data showed that low-participation students usually stopped participating between the third and fourth weeks of the courses, which spanned at least nine weeks. For this reason, much of the data for students in this group showed little change between the midpoint and end of the courses.

Another observation was the large difference in the percentage of students with insufficient performance between the low-participation and high-participation groups. This difference ranged between 27% and 33% for the three courses.

Therefore, the argument that the peer-assessment system could identify the large majority of students who may be at-risk of failing in a timely manner is supported by these observations. The charts presented in Figures 1, 2, and 3 demonstrate the differences in performance levels between low-participation and high-participation groups for the three courses.
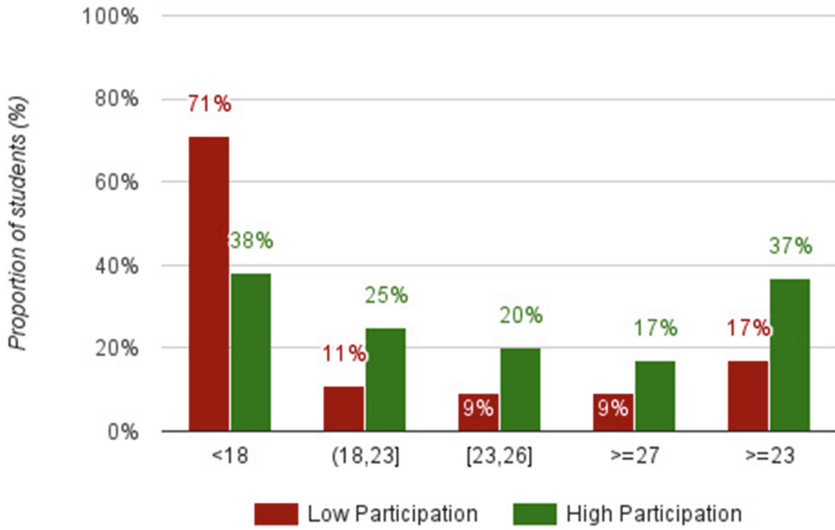
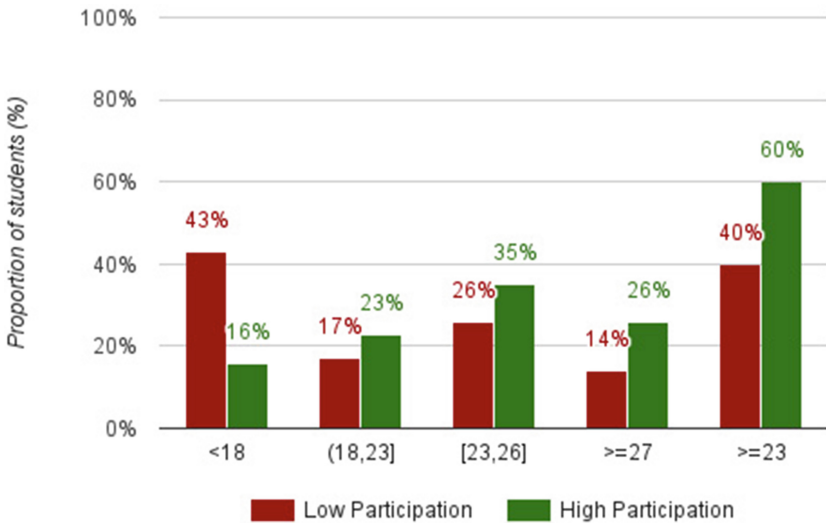**Fig. 1.** Participation in peer-assessment tasks and exam scores for course IG1

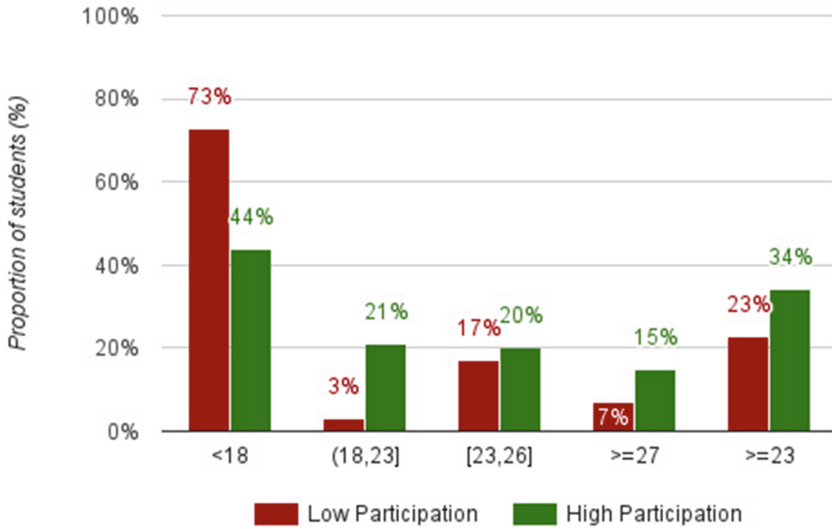**Fig. 2.** Participation in peer-assessment tasks and exam scores for course PR2

Fig. 3. Participation in peer-assessment tasks and exam scores for course LP

5   Discussion and Conclusion

In order to examine if lack of participation in online peer-assessment activities for a course could fairly identify students who would not successfully complete the course, the data of the over 600 students in enrolled in three courses were analysed. The analyses revealed that, after multiple exam sessions, all of which a student can sit, the majority of students with little participation struggled to either pass the exams or perform well. The findings contribute to yet another motivation to automate peer-assessment activities.

Although, at this stage, there is not enough evidence to suggest that participation in online peer-assessment tasks improves overall student performance, the argument that these activities could provide well-timed identification of students who may fall behind at later stages of the course was supported by the findings.

Participation in peer-assessment activities was not mandatory. Nonetheless, the large majority of student had completed at least a third of the tasks, which ranged between 22 and 27 for the three courses. Earlier survey results also showed generally positive reception of the practice among students. Hence, studies of whether the system could promote student engagement or produce learning effects are in order. Moreover, categorisation and mapping of students into more than two performance groups could provide better insights with respect to identifying students that may not be at risk but may still need closer supervision. These two possible extensions of this work will be explored in an upcoming study.

The case for introducing electronic peer-assessment environments into the classroom is supported by the foreseen significant improvements in efficiency and effectiveness of the activities involved. It is hoped that the prospects explored in this study

contribute to the case for transitioning into cost-effective, ubiquitous, and highly interactive electronic peer-assessment solutions.

In particular, the transition to a ubiquitous system can be made with little difficulty by taking advantage of the fact that virtually all students own smartphones or tablets. Developing a mobile peer-assessment solution has the potential to increase student productivity given that peer-assessment tasks are designed to be simple, mobile-friendly and with special attention to privacy and other social aspects. Hence, development of a mobile version of the peer-assessment system will be addressed in the near future.

Calibrating peer-assigned marks as well as training students how to grade their peers' answers have been shown to improve the effectiveness of peer-assessment, with what regards its validity and reliability. Hence, future work will also look to improve the quality of peer-assigned marks by applying these techniques.

References

1. Topping, K.: Peer assessment between students in colleges and universities. Rev. Educ. Res. **68**(3), 249–276 (1998). doi:10.3102/00346543068003249, URL: http://rer.sagepub.com/content/68/3/249.abstract, http://rer.sagepub.com/content/68/3/249.full.pdf+html
2. 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, doi:10.3102/00346543070003287, URL http://rer.sagepub.com/content/70/3/287.abstract, http://rer.sagepub.com/content/70/3/287.full.pdf+html
3. Ashenafi, M.M.: Peer-assessment in higher education twenty-first century practices, challenges and the way forward. Assess. Eval. High. Educ. 1–26 (2015). doi:10.1080/02602938.2015.1100711, URL: http://dx.doi.org/10.1080/02602938.2015.1100711
4. Ashenafi, M.M., Riccardi, G., Ronchetti, M.: Predicting students' final exam scores from their course activities. In: Frontiers in Education Conference (FIE), pp. 1–9, 32614 2015. IEEE. doi:10.1109/FIE.2015.7344081
5. Ashenafi, M.M., Ronchetti, M., Riccardi, G.: Predicting student progress us- ing peer-assessment data. In: 9th International Conference on Educational Data Mining (EDM) (2016)
6. Luxton-Reilly, A.: A systematic review of tools that support peer assessment. Comput. Sci. Educ. **19**(4), 209–232 (2009). doi:10.1080/08993400903384844, URL: http://dx.doi.org/10.1080/08993400903384844, http://dx.doi.org/10.1080/08993400903384844
7. de Raadt, M., Lai, D., Watson, R.: An evaluation of electronic individual peer assessment in an introductory programming course. In: Proceedings of the Seventh Baltic Sea Conference on Computing Education Research (Koli Calling 2007), vol. 88, pp. 53–64. Australian Computer Society, Inc., Darlinghurst, Australia (2007). URL: http://dl.acm.org/citation.cfm?id=2449323.2449330
8. Denny, P., Hamer, J., Luxton-Reilly, A., Purchase, H.: Peerwise: Students sharing their multiple choice questions. In: Proceedings of the Fourth International Work- shop on Computing Education Research (ICER 2008), pp. 51–58. ACM, New York (2008). doi:10.1145/1404520.1404526, URL: http://doi.acm.org/10.1145/1404520.1404526

9. Paré, D., Joordens, S.: Peering into large lectures: examining peer and expert mark agreement using peerscholar, an online peer assessment tool. J. Comput. Assist. Learn. **24**(6), 526–540 (2008). doi:10.1111/j.1365-2729.2008.00290.x, URL: http://dx.doi.org/10.1111/j.1365-2729.2008.00290.x

10. Fagen, A.P., Crouch, C.H., Mazur, E.: Peer instruction: results from a range of classrooms. Phys. Teach. **40**(4), 206–209 (2002). doi:http://dx.doi.org/10.1119/1.1474140, URL: http://scitation.aip.org/content/aapt/journal/tpt/40/4/10.1119/1.1474140

11. Simon, B., Kohanfars, M., Lee, J., Tamayo, K., Cutts, Q.: Experience report: peer instruction in introductory computing. In: Proceedings of the 41st ACM Technical Symposium on Computer Science Education (SIGCSE 2010), pp. 341–345. ACM, New York (2010). doi:10.1145/1734263.1734381, URL: http://doi.acm.org/10.1145/1734263.1734381

12. Kennedy, G.E., Cutts, Q.: The association between students' use of an electronic voting system and their learning outcomes. J. Comput. Assist. Learn. **21**(4), 260–268 (2005). doi:10.1111/j.1365-2729.2005.00133.x, URL: http://dx.doi.org/10.1111/j.1365-2729.2005.00133.x

13. Yadav, R.K., Gehringer, E.F.: Automated met reviewing: a classifier approach to assess the quality of reviews. In: Computer-Supported Peer Review in Education (CSPRED-2016) (2016)

14. Wang, Y., Wang, H., Schunn, C., Baehr, E.: Choosing a better moment to assign reviewers in peer assessment: the earlier the better, or the later the better? In: Computer-Supported Peer Review in Education (CSPRED-2016) (2016)

15. Morris, J., Kidd, J.: Teaching students to give and to receive: improving interdisciplinary writing through peer review. In: Computer-Supported Peer Review in Education (CSPRED-2016) (2016)

16. Song, Y., Gehringer, E.F., Morris, J., Kid, J., Ringleb, S.: Toward better training in peer assessment: does calibration help? In: Computer-Supported Peer Review in Education (CSPRED-2016) (2016)

17. Gehringer, E.F.: A survey of methods for improving review quality. In: Cao, Y., Väljataga, T., Tang, Jeff, K.,T., Leung, H., Laanpere, M. (eds.) ICWL 2014. LNCS, vol. 8699, pp. 92–97. Springer, Heidelberg (2014). doi:10.1007/978-3-319-13296-9_10

18. Babik, D., Gehringer, E.F., Kidd, J., Pramudianto, F., Tinapple, D.: Probing the landscape: Toward a systematic taxonomy of online peer assessment systems in education. In: Computer-Supported Peer Review in Education (CSPRED-2016) (2016)