Feasibility and Acceptability of Peer Assessment for Coding Assignments in Large Lecture Based Programming Engineering Courses

Christine E. King, PhD
Department of Biomedical Engineering
University of California, Irvine
Irvine, CA USA
kingce@uci.edu

***Abstract*—This Research to Practice Full Paper presents the feasibility and student acceptability of peer assessment methods for a large lecture-based computer programming course. Peer assessment is a widely used and successful method for learning in higher education, and has been extensively used in diverse fields. However, the application of this technique in computer programming courses has been limited, and typically has only been feasibly performed in online courses and provided in an online format. For those courses where students attend lectures with large class sizes, in-person peer assessment may provide a more effective technique to learn a programming language while providing the necessary feedback that instructors are not able to provide due to limited resources. As a result, the following research questions were posed: 1) is it feasible to implement in-person peer assessments of coding assignments in a large lecture based course? 2) to what extent do students and instructors find this approach acceptable? 3) do students who find peer grading acceptable improve their homework performance on coding assignments? To test the feasibility of this concept, in-person peer assessments of coding assignments were performed in a large programming course. Specifically, 118 students assessed a paper version of their peer's coding assignment. These sessions were compared to a control assignment that was not peer assessed. Quantitative as well as qualitative assessments of whether the peer assessment sessions improved students' coding abilities were analyzed. It was found that it was feasible to implement in-person peer assessment for large class sizes. From online surveys, it was found that students enjoyed the peer assessment session, and had an overall positive attitude. Finally, the positive findings on the effect of students' abilities to learn a programming language and improve their homework performance demonstrated that this technique may become an important tool to improve student learning in large lecture based programming courses. As a result, this study fills a considerable gap in the literature regarding in-person peer grading for large class sizes. If further tested, this technique may shape future curricula of large programming courses and provide a scalable technique for creating meaningful feedback on assignments.**

Keywords*— *programming language course, coding, peer assessment

I. Introduction

Peer assessment is a widely used and successful method for learning in higher education, and has been extensively used in diverse fields, such as academic writing, business, science, and medicine [1 – 4]. In general, the method requires students in a class to grade an assignment or a performance of another student and provide feedback on the assignment. This allows students to be exposed to different perspectives of others' works, and by providing and receiving feedback, allows students to have a better understanding of the topic [5]. In addition, students learn through seeing the cognitive products and processes of peers' work [6]. Through these advantages of peer grading, it has been found that peer grading is correlated with effective learning [7-9].

Peer grading has been widely implemented across different computer programming courses due to the positive impact on student learning, [10, 11]. However, the application of this technique in these courses has been limited, and typically has only been feasibly performed in online courses and provided in an online format [6, 10-16]. Specifically, there have been many studies that have used web-based peer grading [6, 10-16], but very few studies have examined the feasibility of in-person peer grading for large lecture-based computer programming courses.

As student to staff ratios are increasing, it is important to maintain the quality of assessment and feedback in courses where students attend lectures with large class sizes. In-person peer assessment may provide an effective technique to learn a programming language while providing the necessary feedback that instructors are not able to give due to limited resources. Furthermore, in-person peer assessments may lead to more out of class discussions on programming techniques among students that would not be possible in online based courses where they do not physically meet, which may in turn promote active learning [17, 18].

This study assesses the feasibility and acceptability of in-person peer assessments of coding assignments for large lecture-based computer programming courses. The feasibility of the technique was determined by performing paper-based peer assessment of a homework assignment, where each student was given a peer's homework assignment on different programming techniques to grade and provide feedback. The students' acceptability of this approach was assessed through an online survey, while the instructors' acceptability was assessed through a focus group interview. In addition, student

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

acceptability and its impact on their homework scores was determined by comparing a peer graded assignment to a non-peer graded assignment. The following sections will describe the related prior work as well as the methods and results of the feasibility and acceptability study.

II. Related Work

As previously mentioned, much of the prior work has focused on performing peer grading in an online format [6, 10-16]. For example, the researchers in [6] used an Internet-based peer-assessment system *NetPeas* to perform peer grading on assignments for 58 computer science students. Similarly, researchers in [11] utilized *Expertiza*, an Internet-based peer assessment system, to perform peer assessment on 10 programming assignments and 51 writing assignments across three academic years. These systems, along with many others such as *PRAISE* (Peer Review Assignments Increase Student Experience) [19], *PeerWise* [20], and *PeerScholar* [21], utilize generic assessment tools and rating systems to assess their peers' assignments for different submission types such as multiple choice responses, essays, and code.

Online peer assessment systems provide appropriate administrative support for the management of the peer review process [22]. This has been found to be an important limiting factor in implementing peer grading in large classes [21]. In addition, online tools allow for anonymous online discussion, improved legibility, faster turn-around time, and easy reuse of comments. These benefits of online peer-assessment make these tools attractive, however, they do not provide the socio-affective elements that are difficult to develop in online interactions.

In-person peer assessment systems can provide the same benefits of online peer assessment as well as additional socio-affective elements [23] such as the development of personal and group relationships that have a positive influence on the learning experience. In addition, it has been found that when compared to online peer assessment, in-person peer assessment can improve student response rate [24], as well as provide easier communication between students to provide hints and helpful feedback [25]. In-person peer assessment also allows instructors to be able to perform facilitative coaching and training, and monitor and assess the peer grading process with greater reliability and validity. These benefits in addition to others, such as modelling of competence, enthusiasm and coping, building trust, facilitating disclosure, and developing loyalty and accountability [23], allow in-person peer assessment to have significant benefits for effective learning in the classroom.

Prior research [24] has found that in-person peer assessment requires significant administrative overhead, which is an important limiting factor in applying this technique in large lecture-based classrooms. Examples of administrative overhead include the time taken to explain the peer grading process, organization of the papers, student training, and administration of the assignments across students [24]. To determine if administrative overhead can be minimized to make in-person peer grading feasible in a large-lecture based class, this study assesses the feasibility of performing in-person peer assessment of coding assignments for a large lecture based computer programming course. The burden associated with administrative overhead was determined during the feasibility testing, and the acceptability of the technique to provide feedback on coding assignments was assessed among instructors, teaching assistants, and students.

III. Methods

A. Objective

To assess the feasibility of in-person peer grading for large lecture-based computer programming courses, the following research questions were posed:

1) Is it feasible to implement in-person peer assessments of coding assignments in a large lecture based course?

2) To what extent do students and instructors find this approach acceptable?

3) Do students who find peer grading acceptable improve their homework performance on coding assignments?

While the above research questions are straightforward, they have significant implications in determining how to create a scalable technique for providing meaningful feedback on assignments while maintaining the positive socio-affective elements of performing in-person peer assessment.

B. Recruitment and Participation

Students in a large programming course at an R1 (highest research activity) university performed in-person peer assessment during one of their coding homework assignments. Specifically, 118 students in a sophomore-level engineering analysis course that focuses on Matlab computer programming for biomedical applications were recruited to participate in the study. This course was co-instructed by two instructors and one teaching assistant, and provided three one-hour lectures and one one-hour discussion each week for ten weeks. All students, instructors, and teaching assistant that participated in the study provided informed consent (University of California Irvine IRB Approval Number: 2018-4211).

All students participated in the in-person peer assessment of the homework assignment and submission of the control homework assignment. Prior to completion of the course, students were asked to voluntarily complete a brief online survey to ascertain whether the in-person peer assessment technique was enjoyable, improved their coding abilities, and was useful in learning relevant course learning outcomes. To maximize compliance, completion of the survey awarded the student with one extra credit point towards their final grade. Out of the 118 students that participated in the study, 106 students responded to the online survey. The student demographics are presented in Table 1 below.

TABLE I. DEMOGRAPHICS OF THE STUDY PARTICIPANTS.

Number of Participants: 106	
Low Income	24 (22.64%)
First Generation	29 (27.36%)
Females	55 (51.89%)
Underrepresented Minorities	24 (22.64%)
Year in School based on Number of Credit Units:	
Freshman	0 (0.00%)
Sophomore	22 (20.75%)
Junior	65 (61.32%)
Senior	19 (17.92%)

In addition to the student online survey, interviews among the instructors and teaching assistant was also performed. These interviews were designed to examine the feasibility of the approach, as well as its benefits, pitfalls, and potential improvements for applying the technique in large lecture-based computer programming courses.

C. Feasibility Testing

To test the feasibility of this concept, in-person peer assessments of coding assignments were performed during one of the assignments and was compared to a subsequent control assignment in which no peer assessment was performed. Specifically, students assessed a paper version of their peer's coding assignment during their in-person discussion section. The discussion sections grouped students into classes of 50, which allowed the instructors and teaching assistants to easily hand out the students' paper-based homework assignment to their peers. This helped reduce the administrative overhead associated with organizing the peer assessment session.

To peer grade the assignment, students were asked to provide constructive feedback in terms of correctness, clarity of the code, and unique or alternative methods for algorithm design. They were told that the feedback provided by their peers had no impact on their assignment grade, and the names of both the grader and assessor were not anonymized in hopes of creating a sharing of knowledge with their peers. After the assessments, students were asked to find unique algorithm designs and present them to the entire class. The instructors and teaching assistants then provided a think-pair-share session in which the students in groups of 6-8 thought about a unique algorithm design that differed across their peer's assignments to present. If a unique concept was found, the instructor or teaching assistant called on that group to each present their unique solution, and to comment on the diversity of the solutions and techniques employed by their peers to achieve the same result. This was performed as to encourage group discussion of algorithm design concepts, and to elicit discussions outside the classroom of the concepts related to the assignment.

The peer grading sessions were compared to a control assignment that was not peer assessed. Discussion sections after the control assignment did not include peer grading, rather, instructors and teaching assistants performed traditional in class examples on related concepts discussed during lecture. Feedback on the feasibility and logistics of peer grading homework assignments for the course was assessed through in person interviews with the instructors and teaching assistants.

D. Acceptability Asesssments

Quantitative as well as qualitative assessments of whether the peer assessment sessions improved students' coding abilities were analyzed. To this end, a brief online survey was provided to the students after completion of both the peer assessment assignment and control non-peer assessed. This was performed to assess the acceptability of the technique from the student's perspective. As presented in Table 2, the following constructs were examined using the survey.

TABLE II. CONSTRUCTS AND RESULTING SURVEY QUESTIONS TO ASSESS ACCEPTABILITY OF PEER ASSESSMENT FOR PROGRAMMING HOMEWORK ASSIGNMENTS.

Construct	**Question**
Acceptability of peer grading	Did you enjoy performing peer grading on Homework 1?
Positive nature of feedback	What did you like about peer grading Homework 1?
Negative nature of feedback	What did you not like about peer grading Homework 1?
Perceived utility of peer grading on learning outcomes	Did you wish we did peer grading on Homework 2 if there had been time?
Perceived utility of peer grading on coding ability	Did peer grading help improve your coding abilities?

As seen in Table 2, students' perceptions of the experience were ascertained using both 3-point categorical items and open-ended questions. Both the positive and negative nature of the peer assessment session were assessed so as to determine which key factors led to student acceptability and which factors were problematic or limited the acceptability of the technique. In addition, perceived usefulness and utility of the peer assessment session was determined by comparing the peer assessment session to the controlled non-peer assessed session, and asking whether this form of assessment had improved their programming skills. These questions were proposed in hopes of identifying the strengths, weaknesses, and perceived effectiveness of the technique among students.

To determine if the survey responses such as student acceptability were dependent upon categorical demographic variables such as sex, low income, and race, statistical analysis was performed. Specifically, through cross tabulation, Pearson chi-square tests were performed to determine whether the categorical variables were associated or independent.

Other variables that were quantitatively explored in terms of their relationship with student acceptability included ordinal variables such as year in school based on number of credit units. For these variables, Pearson chi-square tests through cross tabulation were also performed. It can be noted that the year in school variable contained sophomore, junior, and senior class levels.

To compare student acceptability with the student's final grade, the Kruskal Wallis test was performed due to the fact that the assumptions required for ANOVA were violated. Particularly, this statistical test was chosen to examine the final grade against the three categories of student acceptability as it was found that the final grade scores had a bimodal and negative skew distribution type, and thus the variable could not be considered as Gaussian. As a result, the Kruskal Wallis test was chosen as it is a non-parametric method of ANOVA. It is also the generalized form of the Mann-Whitney test, as it permits two or more groups to be compared for statistical significant difference.

In order to determine whether there was an improvement in homework score before and after peer grading, multinomial regression for categorical data was performed to compare the effect of acceptability on the difference in homework score. Note that the homework assignments were graded on a check-minus, check, and check-plus ordinal scale with a 10% deduction for late submissions, so the scores were considered on an ordinal scale rather than a continuous one.

In addition to assessing student acceptability of peer-assessment, strengths, weaknesses, and recommended improvements to the procedure from an instructor's perspective were obtained through a focus group. The following open-ended questions were asked to evaluate the experience of performing peer assessment in a large lecture-based classroom:

1) Was peer grading in person feasible to perform for homework 1?
2) What did you like about peer grading?
3) What did you not like about peer grading?
4) What would you do the same if you were to try peer grading again in the classroom?
5) What would you do differently if you were to try peer grading again in the classroom?

It can be seen from the above interview questions that similar to the student online survey, both the positive and negative nature of the peer assessment session were analyzed. In addition, the factors that were successful in implementing the technique as well as recommended improvements were evaluated from an instructor's perspective through questions 4 and 5. These questions were proposed to complement the findings from the student online survey, and similarly to the survey, identify the strengths, weaknesses, and effectiveness of the technique among instructors.

IV. Results

It was found that it was feasible to implement in-person peer assessments of assignments for large class sizes. In addition, the post peer assessment discussion demonstrated that an open dialog of the complexity of problems and alternative solutions helped encourage students to think "outside the box". From online surveys, it was found that students enjoyed the open feedback and group discussions, and had an overall positive attitude.

A. Feasibility Testing Results

After performing the instructor and teaching assistant interviews, it was found that peer grading homework assignments in the large computer programming course was feasible. Specifically, the interviews revealed that the peer grading activities alleviated much of the burden associated with providing meaningful feedback to students on their assignments given limited time resources.

When asked about the strengths of the peer grading session, the instructors and teaching assistant similarly found the ability for students to learn from each other highly beneficial, as it created social and cognitive interaction. This is an important benefit for in person peer grading over online and anonymous peer grading techniques, as social and cognitive interaction with peers is important in enhancing active participation and learning [27]. In particular, the instructors and teaching assistant mentioned that:

Instructor 1: "*I believe that students can learn from each other with effective, judicious use of peer grading. They are able to see real examples of how their peers approach writing code.*"

Instructor 2: "*By having students provide feedback on their peers' assignments, I now have time to develop more active learning techniques for lectures, such as games, clicker questions, and group work. In addition, I like that students can explore various ways of problem solving, as oftentimes in engineering and algorithm design, there are multiple solutions to the same problem.*"

Teaching Assistant: *"I think students were more socially active during the peer grading process."*

In addition to assessing the strengths of the peer grading session, the weaknesses were also examined. The instructors and teaching assistant all agreed that the implementation of the peer grading technique on the homework was not as effective as it could have been. They found that although students were more inclined to perform outside class discussions and create a sharing of knowledge, the lack of anonymity was also seen as a negative aspect of the process. In addition, the instructors mentioned that the discussions on unique algorithm designs and alternative solutions that resulted from the peer grading session should have been better integrated into the in-class lectures.

The instructors and teaching assistant found that the implementation of the peer grading technique was overall successful, and that the session actively engaged the students to discuss and understand other students' methods to accomplish the same task. However, to further improve the success of this technique, one instructor suggested that the students should be:

"*…required to do this at home – take someone else's code, interpret it, and write up a short report that involves a summary of the student's evaluation and also a reflective* *statement on what they learned from the process."* (Instructor 1)

The other instructor also suggested that more structure in the peer grading process would have made the technique more successful. However, instead of performing peer grading at home and providing a summary on the evaluation, the instructor mentioned that providing a training session and rubric on how to peer grade prior to starting their own assignments would result in better feedback across students. The implementation of a calibrated peer assessment, however, would increase the amount administrative overhead and may be difficult to implement in such a large class size. Finally, the teaching assistant also mentioned that more structure during the peer grading sessions would make them more effective, suggesting to provide extra credit points to those who had a similar evaluation to the expert (i.e. the teaching assistant's evaluation of the assignment).

B. Acceptability Testing Results

The acceptability questions of the instructor interviews and student surveys revealed that the peer grading session was very beneficial for providing meaningful feedback on students' coding assignments and an open discussion about unique algorithm designs. Specifically, after analyzing the student survey responses, it was found that only 8% of the student did not enjoy the peer grading session (Table 3).

TABLE III. TOTAL STUDENT RESPONSESTO THE CATEGORICAL QUESTIONS IN THE ONLINE SURVEY.

Did you enjoy performing peer grading on Homework 1?	
I did not enjoy it	8.49% (9)
I enjoyed it somewhat	64.15% (68)
I enjoyed it a lot	27.36% (29)
Did you wish we did peer grading on Homework 2 if there had been time?	
No	40.57%
Yes	59.43%
Did peer grading help improve your coding abilities?	
No	37.74%
Yes	62.26%

Among the 106 students who responded to the online survey, the most common response about the positive nature of the peer grading session was that the students were able to see and discuss alternate methods to create the same result. Specifically, the following positive comments highlight this benefit of peer grading in a large lecture based computer programming class:

"I liked that I was able to see different strategies to code the same problem. It was really cool to see people's takes on it and see how I could improve my own work (or realize mistakes I made). It also forces me to think about the assignment in much more depth for a second time. By forcing me to review the material again, I get a better understanding of what's expected of me."

"I do like seeing how other people code. Since everyone's background is different, there are many ways to attack a problem that I didn't think about."

"It was very useful to see different people use different ways to complete the same tasks. Also to have the other person explain it to you of how they did it, one-on-one, made it much easier to understand."

It can be seen from the above selected comments that students enjoyed the process of seeing other student's code and their approaches to the same problem. They also enjoyed the learning process of the peer grading session, as many students were inspired to use their feedback to improve their own coding abilities, a benefit that would not be otherwise possible given the burden associated with instructors and teaching assistants being able to provide meaningful feedback to students given their limited time resources.

Other positive findings of the peer grading session were that 62% of the students found that the peer grading session helped improve their coding abilities (Table 3). In particular, several students mentioned in the question on the positive nature of the peer grading session that they were able to improve their own coding abilities by being able to see how their peers were able to accomplish the same task. For example, as seen in the quotes below, the following students believed that the peer grading session allowed them to be able to improve their coding abilities both through the open discussion session as well as through the formal feedback they received on their own assignment by their peers. In addition to improving their coding skills, it can be noted by the last two quotes that some students believed that the peer grading session improved their overall engineering problem solving skills.

"It was really interesting to see, other peoples thought process. I remember it took me like 10 lines of code to accomplish a certain step and with the peer grading I saw how one of my peers accomplished the same thing in less than 5 lines. That in turn, inspired me to become more creative with my future codes. I also got really good feedback from my classmate that commented on my code."

"I can see the different possibilities on how to go about the certain problems, especially the first one (homework problem). I also saw the flaws in my code that I didn't beforehand due. I believe it really does help the learning process."

"It help me to find more ways to figure out questions."

"Gaining a different perspective on how to approach problems. I have a hard time starting problems but once I start I usually can finish."

The assessment of the perceived utility of peer grading on other learning outcomes revealed that only 59% of the students found utility in the technique and would have preferred performing peer grading on the control non-peer graded homework assignment (Table 3). This may be due to the some of the perceived weaknesses of the peer grading session. For instance, some students found that some of the administrative overhead to perform the session was still too long, even after performing the peer grading sessions during the in-person discussion sections where students were grouped into classes of 50, and further breaking the students down into groups of 6-8 students. The main issue was not due to the grouping of students, but rather due to how the assignments were distributed within the in-person discussion section. Namely, students found that distributing paper-based assignments across 50 students took too long. Two of the students even mentioned this directly, and explained that the distribution of the peer grading assignment should be further limited to only those in the smaller group sizes of 6-8 students:

"How we distributed the homework, it would have been easier to switch with people at our table."

"…I also think it'd be good to do peer grading by table instead of randomly distributing it so we could get our paper back by the end of discussion and ask each other questions or have discussions on how we all approached the problem!"

Other weaknesses of the peer grading session also contributed to the perceived utility of the technique. In particular, some students found that the writing portion of the peer grading session was time inefficient, even suggesting that:

"Rather than writing feedback, I prefer talking one-on-one to receive direct feedback while exchanging thoughts on the homework."

In addition to suggestions for improving the mode of providing feedback to further reduce time inefficiencies, some students did not like the criteria for peer grading. This refers to how the peer grading session was introduced, and how they were asked to provide constructive feedback in terms of correctness, clarity of the code, and unique or alternative methods for algorithm design. As seen below, one student even suggested providing a rubric at each table to help guide them during the peer grading session. This aligns well with the reflections from one of the instructors that a calibration session would improve the success of the technique.

"The criteria for peer grading was a bit vague at first. It would be nice to have a super general rubric or something that would guide us on specific things to look for. With the large variety of solutions, it might be hard to do though. I also think it'd be cool if we asked people to make sure their outputs were also printed when peer grading because it's hard to keep up with what certain people print out if there's a lot of outputs."

Finally, other issues associated with the peer grading session that may have reduced perceived utility of the technique included the lack of anonymity of the assignment. For example, one student found that:

"Peer grading added a sort of pressure, where I felt that other people would realize that my coding skills are not at the same level as everybody else. However, as I was peer grading, I realized that we all had areas of improvement, simply from talking to those at the same table as myself."

This statement describes both the benefits as well as pitfalls of the purposeful lack of anonymity in the peer grading session. Originally, the peer grading session was designed to include names of both the grader and assessor in hopes of creating a sharing of knowledge and outside class discussion among students, as well as reduce administrative time. In addition to these benefits, the lack of anonymity also improved student understanding and self-confidence. This is consistent with prior findings on peer assessment in higher education [28 – 30].

Even though more than half of the students perceived that there is utility in performing peer grading on their computer programming assignments to improve learning outcomes, the statistical analyses that compared student performance on the homework assignments to their acceptability of the peer grading session revealed that the peer grading session improved student performance.

A Pearson's chi-square test was performed to determine whether student acceptability of peer grading is correlated with students' homework score level. To this end, after comparing the students' performance on the peer-graded assignment to the control non-peer graded assignment (Figure 4), it is clear that students had a higher overall performance on the control non-peer graded assignment than on the peer-graded one. This is because the control homework assignment was performed *after* the peer graded assignment, so the effect of the students' ability to learn from the peer grading session can be seen by the increased performances of the second control homework assignment. By analyzing the homework performances, it is clear that the students who enjoyed peer grading improved their homework performance level (Table 4). Furthermore, the multinomial regression analysis that assessed the relationship between student acceptability level and improvement in homework score level found that there is a statistically significant association between improvement in student homework performance after peer grading and the acceptability level of the peer grading technique (p-value: 0.002, standard error: 0.219).

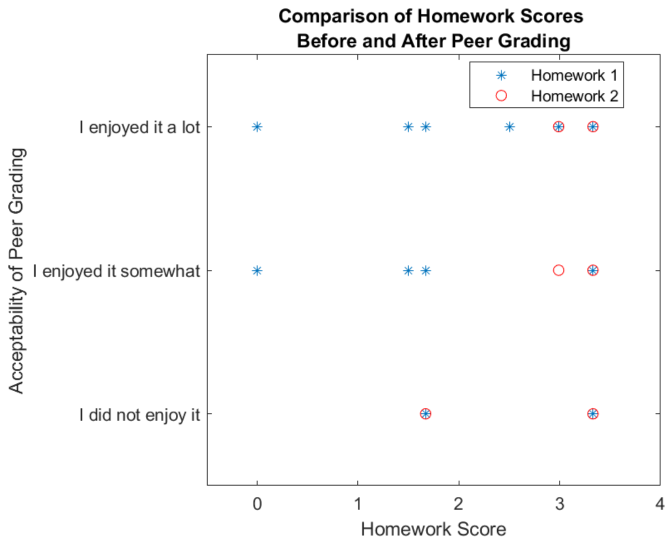

Fig. 1. Comparison of homework scores before and after the peer grading session. Note that Homework 1 was the peer grading session assignment, and Homework 2 was the control non-peer graded assignment.

TABLE IV. PROPORTIONS OF STUDENT HOMEWORK SCORE IMPROVEMENT COMPARED TO ACCEPTABILITY OF PEER GRADING TECHNIQUE. NOTE THAT THE SCORE LABEL IS BASED ON A 3 POINT ORDINAL SCALE WITH WITH A 10% DEDUCTION FOR LATE SUBMISSIONS.

Home work	Score Label	Did Not Enjoy Peer Grading	Somewhat Enjoyed Peer Grading	Enjoyed Peer Grading
1	0	0	2	1
	1.5	0	1	1
	1.67	1	4	6
	2.5	0	0	1
	2.99	0	0	1
	3.33	8	61	19
2	0	0	1	0
	1.5	0	0	0
	1.67	1	0	0
	2.5	0	0	0
	2.99	0	5	2
	3.33	8	62	27

The Pearson chi-square tests performed on other variables such as demographics (sex, low income, race), academic level (year in school based on number of credit units) did not find any statistically significant relationship with student acceptability level of the peer grading session. Furthermore, the Kruskal Wallis test performed to compare student acceptability with the student's final grade did not find a statistically significant relationship. These results show that there were no demographic or academic level biases towards student acceptability of peer grading. This is important, as it demonstrates that the peer grading session was perceived equally acceptable among diverse groups.

V. DISCUSSION

The results of this study indicate that peer grading is both feasible and acceptable among students and instructors for large lecture based computer programming courses. The peer grading session performed was feasible in a large class size by minimizing administrative overhead using smaller group discussion sections, which is scalable to larger lecture based classes. In addition, the peer grading session improved student performance on their post-session homework assignment, suggesting that utilizing peer grading in these types of classes can improve student learning.

The positive findings mentioned above lead to several recommendations for performing peer grading in large lecture-based computer programming courses. First, instructors that are considering using this technique must decide whether to preserve anonymity in the peer grading session. As mentioned in prior studies [26, 32], preserving anonymity can reduce the opportunity for collusion and cheating, remove bias in terms of friendship and enmity effect, and lessen the difficulty and embarrassment of marking friends' work. However, creating anonymity will also increase the administration time due to the distribution systems needed to track the peer grading process while maintaining author and marker confidentiality [26]. Creating anonymity will also remove the positive social and cognitive interaction with peers, an important aspect of the peer grading technique to enhance outside class peer-to-peer discussion and learning. The instructors who performed the peer grading session in this study decided that it was more important to reduce administrative time constraints and improve outside class student discussion, so the peer grading session were purposefully performed with student names. This proved to be beneficial, as the results from the study indicate that this decision was an important aspect to the success of the technique.

It is recommended that in addition to providing student names, the technique could be further improved upon by having students perform a short one-on-one discussion of their assignment with the peer grader immediately after the peer grading session to further encourage outside class discussion. Given the time constraints already associated with administration of peer grading in large lecture based classes, however, this discussion may be limited to only a few minutes and should be performed at the very end of the class.

To improve inefficiencies associated with administrative overhead such as the time required to distribute the peer grading assignment, it is suggested to break students into sub-groups. This allows for students to easily randomize and distribute their assignments, and encourages small group discussion. Furthermore, by requiring students to bring a paper-based copy of their coding assignment and distribute it randomly within their own sub-group, the administrative task is now assigned to the students rather than the instructors.

Finally, the delivery of the explanation of the peer grading process can be improved by performing calibration outside the classroom, similar to the Calibrated Peer Review method used in reading and writing applications [32]. This could include using an outside class assignment that requires students to submit a short code on a specific topic, and then review the "calibration" code for different performance levels under clear grading criteria to train the students as reviewers. By providing a very clear criteria for the peer grading process for coding assignments, students will be better prepared for the peer grading session and instructors will not need to perform an explanation of the process in class, further reducing administration time.

VI. CONCLUSION

The positive findings and their effect on students' abilities to learn a programming language demonstrated that this innovative technique for in-person peer grading in large lecture based programming courses may become an important tool to improve student learning. If further tested and improved upon given the above recommendations, this technique will have a more significant impact on lecture based computer programming courses than online peer assessment. Specifically, as lecture-based computer programming courses are increasingly enlarging in class size, the resources available to instructors and staff to provide meaningful feedback on homework assignments are decreasing. In-person peer assessment in these types of classes can provide a scalable technique for creating meaningful feedback on assignments while improving socio-affective elements, and can potentially shape future curricula of large programming courses.

ACKNOWLEDGEMENTS

We would like to thank the Division of Teaching Excellence and Innovation for their efforts in assisting with IRB approval, and Dr. Kameryn Denaro at the Teaching and Learning Research Center at the University of California Irvine for her assistance with demographic data collection. We would also like to thank the students, instructors, and teaching assistant of the course for their participation in the study.

REFERENCES

[1] N. Falchikov, "Peer feedback marking: Developing peer assessment," *Innovasdfadsfations in Education and Training International*, vol. 32, pp. 175–187, 1995.

[2] M. Freeman, "Peer assessment by groups of group work," *Assessment and Evaluation in Higher Education,* vol. 20, pp. 289–300, 1995.

[3] I.B. Strachan, and S. Wilcox, "Peer and self assessment of group work: developing an effective response to increased enrollment in a third year course in microclimatology," *Journal of Geography in Higher Education*, vol. 20, no. 3, pp. 343–353, 1996.

[4] R. Rada, "Efficiency and effectiveness in computer-supported peer-peer learning," *Computers and Education*, vol. 30, no. 3/4, pp. 137–146, 1998.

[5] N. Falchikov and J. Goldfinch, "Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks," *Review of Educational Research*, vol. 70, no. 3, pp. 287-322, 2000.

[6] S.S. Lin, E.Z.F. Liu, and S.M. Yuan, "Web‐based peer assessment: feedback for students with various thinking‐styles," *Journal of Computer Assisted Learning,* vol. 17, no. 4, pp. 420-432, 2001.

[7] T.J. Crooks, "The impact of classroom evaluation practices on students," *Review of Educational Research*, vol. 58, pp. 45-56, 1988.

[8] J.A. Kulik and C.L.C Kulik, "Timing of feedback and verbal learning," *Review of Educational Research*, vol. 58, pp. 79-97, 1988.

[9] R.L. Bangert-Drowns, C.L.C Kulick, J.A. Kulick, M.T. Morgan, "The instructional effect of feedback in test-like events," *Review of Educational Research*, vol. 61, pp. 213-238, 1991.

[10] M.M Ashenafi, M. Ronchetti, and G. Riccardi, "Exploring the role of online peer-assessment as a tool of early intervention." In *International Symposium on Emerging Technologies for Education*, pp. 635-644, October 2016.

[11] Y. Song, Z. Hu, and E.F. Gehringer, "Closing the circle: use of students' responses for peer-assessment rubric improvement," In *International Conference on Web-Based Learning*, pp. 27-36, November 2015.

[12] T. Downing, and I. Brown, "Learning by cooperative publishing on the World Wide Web," *Active Learning*, vol. 7, pp. 14–16, 1997.

[13] R. Davies, and T. Berrow, "An evaluation of the use of computer supported peer review for developing higher-level skills," *Computers & Education*, vol. 30, no. 1-2, pp. 111-115, 1998.

[14] Y. Zhao, "The effects of anonymity on computer-mediated peer review," *International Journal of Educational Telecommunications*, vol. 4, no. 4, pp. 311-345, 1998.

[15] E.F. Gehringer, "Electronic peer review and peer grading in computer-science courses," *ACM SIGCSE Bulletin*, vol. 33, no. 1, pp. 139-143, 2001.

[16] J. Sitthiworachart, and M. Joy, "Effective peer assessment for learning computer programming," In *ACM SIGCSE Bulletin*, vol. 36, no. 3, pp. 122-126, June 2004.

[17] S. Freeman, S.L. Eddy, M. McDonough, M.K. Smith, N. Okoroafor, H. Jordt, and M.P. Wenderoth, "Active learning increases student performance in science, engineering, and mathematics," In *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8410-8415, 2014.

[18] D. Nicol, A. Thomson, and C. Breslin, "Rethinking feedback practices in higher education: a peer review perspective," *Assessment & Evaluation in Higher Education*, vol. 39, no. 1, pp. 102-122, 2014.

[19] M. De Raadt, D. Lai, and R. Watson, "An evaluation of electronic individual peer assessment in an introductory programming course," In *Proceedings of the Seventh Baltic Sea Conference on Computing Education Research*, vol. 88, pp. 53-64, November 2007.

[20] P. Denny, J. Hamer, A. Luxton-Reilly, and H. Purchase, "PeerWise: students sharing their multiple choice questions," In *Proceedings of the Fourth International Workshop on Computing Education Research*, pp. 51-58, September 2008.

[21] D.E. Paré, and S. Joordens, "Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool," *Journal of Computer Assisted Learning*, vol. 24, no. 6, pp. 526-540, 2008.

[22] A. Luxton-Reilly, "A systematic review of tools that support peer assessment," *Computer Science Education*, vol. 19, no. 4, pp. 209-232, 2009.

[23] J. McLuckie, and K.J. Topping. "Transferable skills for online peer learning," *Assessment & Evaluation in Higher Education*, vol. 29, no. 5, pp. 563-584, 2004.

[24] L. Murphy, and D. Wolff, "Take a minute to complete the loop: using electronic Classroom Assessment Techniques in computer science labs," *Journal of Computing Sciences in Colleges*, vol. 21, no. 1, pp. 150-159, 2005.

[25] K. Figl, C. Bauer, and J. Mangler, "Online versus face-to-face peer team reviews," In *36th Annual IEEE Frontiers in Education Conference*, pp. 7-12, October 2006.

[26] R. Ballantyne, K. Hughes, and A. Mylonas, "Developing procedures for implementing peer assessment in large classes using an action research process," *Assessment & Evaluation in Higher Education*, vol. 27, no. 5, pp. 427-441, 2002.

[27] I. Jung, S. Choi, C. Lim, and J. Leem, "Effects of different types of interaction on learning achievement, satisfaction and participation in

web-based instruction," *Innovations in Education and Teaching International*, vol. 39, no. 2, pp. 153-162, 2002.

[28] F.J.R.C. Dochy, M. Segers, and D. Sluijsmans, D, "The use of self-, peer and co-assessment in higher education: A review," *Studies in Higher Education*, vol. 24, no. 3, pp. 331-350, 1999.

[29] G. Mowl, and R. Pain, "Using self and peer assessment to improve students' essay writing: A case study from geography," *Programmed Learning*, vol. 32, no. 4, pp. 324-335, 1995.

[30] K.J. Topping, E.F. Smith, I. Swanson, and A. Elliot, "Formative peer assessment of academic writing between postgraduate students," *Assessment & Evaluation in Higher Education*, vol. 25, no. 2, pp. 149-169, 2000.

[31] P. Davies, "Computerized peer assessment," *Innovations in Education and Training International*, vol. 37, no. 4, pp. 346-355, 2000.

[32] R. Robinson, "Calibrated Peer Review™: an application to increase student reading & writing skills," *The American Biology Teacher*, vol. 63, no. 7, pp. 474-480, 2001.