Learning Together: A Case Study in Applying Peer Learning to a Computer Science Curriculum

Murat K. Gungor, Wren Powell

Abstract

Peer learning is a strategy many educators have turned to in order to add value to the instruction their students receive. Learning from peers has unique advantages and disadvantages compared to learning from a professional instructor, and it is often used as a supplement to formal instruction in the classroom. The success of peer learning is heavily dependent on how it is implemented, and this leaves educators with a need for guidance on how to implement peer learning in their classroom to best meet their goals. There is already a significant body of literature on the topic of peer learning, however heavy emphasis is placed on collaborative writing courses which were designed with peer learning in mind. Writing represents only one of many academic disciplines peer learning could potentially be applied to, and rewriting a curriculum from the ground up around peer learning is a non-starter for many educators. This paper is a case study on the application of peer learning to an existing computer science curriculum in which students are tasked with writing computer code and evaluating their peers' work. In this paper, we provide a review of existing literature on the topic of peer learning and peer assessment, design a system of peer learning which is meant to supplement an existing curriculum and which is informed by existing research in the space, assess the effectiveness of our system using both qualitative and quantitative metrics, and provide recommendations for educators looking to implement a system of peer learning in their classroom.

Introduction

Peer learning is a strategy in education which has seen increasing prevalence in recent years, especially in higher education. It can be defined as "the acquisition of knowledge and skill through active helping and supporting among status equals or matched companions" [1]. Peer learning is "qualitatively different" than learning from a professional instructor and has different advantages and disadvantages. Peer learning has been found to be a cost-effective strategy for learning, providing significant benefit to students with a comparatively low implementation cost on the part of educators, making it an enticing option for educators looking to improve the quality of the education their students receive [2]. When peer learning is implemented with consideration for what specific practices and strategies will best suit the students and the goals of the course, the benefits for students tend to be quite significant [3]. The specific form of peer learning which has seen the most growth and deployment in higher education is peer

assessment, which involves “peers evaluating the products or outcomes of learning of others in the group” [1]. Peer assessment has also been defined 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” [4]. In summary, peer assessment typically involves students assessing the work of their peers in place of or in addition to assessment by an instructor.

There are many elements which characterize a system for peer assessment, and there are many different ways in which such a system can be implemented. Peer assessment is used in a wide variety of subject areas and disciplines, and it is commonly employed with the goals of saving staff time assessing their students’ work or adding value to the instruction students receive. Peer assessment can be applied to a wide variety of types of work, including tests, written work, presentations, and other “skilled professional behaviors.” Peer assessment can act as a substitute for or a supplement to assessment from instructors, it may or may not contribute to a student’s official grade, it can be unidirectional, reciprocal, or mutual, it can be provided anonymously or non-anonymously, it can be provided face-to-face or in a written form, it can occur between students in the same year of study or different years of study, it can be implemented in such a way that more expert students are paired with less expert students or in such a way that students of a similar skill level are paired, it can take place in or out of the classroom, and it is typically implemented such that providing peer assessment is required by staff [4].

By directly involving students in the learning process, peer assessment can promote a sense of “ownership, personal responsibility, and motivation” among students [4]. Peer assessment can also promote social skills which are transferable outside of an educational setting, such as communication, negotiation, diplomacy, empathy, and self-confidence. When assessing the reliability of peer assessment when compared to assessment from an instructor, Topping found that peer assessment is of “adequate reliability and validity in a wide variety of applications” and that it is generally more reliable than self-assessment [4].

Prior research in the space of peer learning and peer assessment can provide insight into what specific practices are most beneficial when implementing peer learning in the classroom. Peer assessment can involve qualitative feedback such as suggestions for improvement or quantitative feedback such as grades or marks. Topping found that providing qualitative feedback rather than grades or marks tends to be more socially comfortable for the assessor and useful for the assessee, however, and it avoids an issue in peer assessment in which assessors exhibit a central tendency in their grading—grading all their peers as “average” [1]. When designing a system of peer assessment, it’s important to ensure that students are provided with instructor assessment as well as peer assessment. Van den Berg et al. found that when peer assessment is used as a substitute for instructor assessment, the total quality of feedback students receive may be hampered [5]. Xu et al. encourages the use of structured peer assessment in which students are prompted with specific questions to answer and specific topics to address in their feedback, and they found that this structured feedback was more “diverse and detailed” than the free-form feedback they received [6].

Instructors often use software to facilitate peer assessment between students. While some software used for peer assessment provides novel features such as allowing students to vote, assign points, and award badges, Melville found that students responded neutrally or negatively to efforts to gamify the peer assessment system in this way, summarizing such

efforts as, “ineffective at achieving enhanced learning motivation and learning experiences” [7]. When analyzing software used for facilitating peer assessment, Melville identified many-to-many threaded commenting as the most compelling feature of the software studied [7]. Van der pol et al. discovered that when students were required to use a software system which did not support threaded comments, many found creative solutions to circumvent that limitation and reply to their peers’ feedback [8].

Of the existing studies in the peer assessment space, some chose to have students take a more collaborative approach to the assignments than just providing feedback. Van den Berg et al. distinguishes between peer assessment and collaborative writing, explaining that the two different approaches can lead to different outcomes [5]. Xu et al. focused on using peer assessment as part of an iterative design process for a graphic design class, where the purpose of the feedback was to help students understand the gap between what message they intended to convey with their work and how audiences actually perceive it [6]. In some studies in this area, students were permitted to see their peers’ submissions before submitting their own or to resubmit their assignment after seeing their peers’ submissions and feedback [5], [7]. As van den Berg et al. notes, there are advantages to taking a more collaborative approach to peer assessment; a shortcoming of simple peer assessment compared to a collaborative writing approach is that peer assessment tends to be restricted to “a simple pattern of (unilateral) comments and responses, in which feedback is indeed provided and accepted (or not, as the case may be), but not discussed” [5]. Saunders also concludes that peer assessment is often a more limited and less effective form of collaborative learning when compared with more interactive methods such as collaborative writing and collaborative editing [9]. The usefulness of peer assessment hinges not just on the quality of the feedback provided, but on an accurate understanding of the provided feedback and its incorporation into the work being assessed. Some of this responsibility falls on the student receiving feedback to guide the dialogue and ask for clarification when necessary [10], [11].

While there are advantages to using peer assessment in higher education, its success is often contingent on how it is implemented. Systems for peer assessment must be adequately organized in order to generate feedback of sufficient quality to be useful to students [5]. Our research provides a case study in how a system of peer assessment was applied to an existing course curriculum and how the decisions made in designing this system of peer assessment influenced educational outcomes. We explain the decisions made in deciding how to incorporate peer assessment into the classroom, the rationale behind those decisions, and we provide a qualitative and quantitative assessment of the success of the experiment. We hope to leave educators with meaningful suggestions for how to incorporate peer learning into an existing curriculum.

Methodology

The main objective of the peer learning system implemented in this study was to expose students to a variety of alternative solutions to the problems they were assigned in order to broaden their understanding of the ways in which a problem can be approached, allow more experienced students to share their expertise, and allow students to learn from their peers’ mistakes. Additionally, the peer learning system implemented for this study was designed with

the goal of minimizing the changes which needed to be made to the curricula of the classes involved; instead of designing a new curriculum around peer learning and peer assessment, an existing curriculum was adapted to make use of peer learning. Minimizing changes to the curriculum also meant minimizing the additional burden put on students and instructors; providing and reviewing peer assessment was not meant to significantly intrude on the time students had for completing assignments or the time instructors had for grading them. This section describes the peer learning system which was designed to meet these criteria and the methodology used by this study to assess its effectiveness.

This research analyzes data from eight separate introductory and intermediate programming classes taught between the Fall of 2018 and the Fall of 2020 as part of an undergraduate computer science and game programming program. These classes taught basic and intermediate programming concepts in C++ and required students to complete several programming assignments over the course of the semester which involved a peer learning component. For each assignment, students were required to review a classmate's code and leave at least one comment providing feedback. Students were graded using a pass/fail system where they were given a passing grade for both submitting a working solution to the problem and leaving a comment on a fellow student's submission providing feedback. The quantity and quality of the feedback students provided was not graded.

Students were not tasked with grading their classmates, and peer assessment did not factor into the final grade for the assignment. Students were able to see all their peers' submissions only once they had submitted the assignment, and were able to leave comments on any of them. However, students were instructed to prioritize providing feedback to students who had not yet received any.

The software used for sharing assignments and providing peer assessment was the Canvas Learning Management System (LMS). This software was already in use by professors for posting assignments, receiving submissions, and posting grades. The software provides a feature which allows students to participate in threaded discussions with other members of their class and allows for locking access to the discussion until they have posted their submission for the assignment. Students were permitted to share their solutions as links to online code editors or as attached archive files. The professor was able to leave individual feedback on student submissions which was not able to be viewed by other students. Canvas LMS was primarily chosen for providing peer assessment due to its current use within the institution. Prior research indicates that the ability for the software to facilitate threaded discussions, where students can reply to other students' comments, could be of significant value. Prior research also indicates that more novel features such as voting and awarding badges would not provide additional value.

In order to translate the peer assessment provided by students into quantitative data, a classification system was formed to categorize the different types of feedback provided and quantify the amount of feedback provided by each comment. Other research in this space has used similar techniques to classify peer feedback; van den Berg et al. created a system which codifies feedback both by the function and the subject of the feedback. This study, however, focused specifically on collaborative writing courses, and we determined that such a system would not translate well to a course focused on reviewing computer code [5].

Our system classifies feedback into three categories: *suggestions*, *compliments*, and *criticisms*; we separate comments into *feedback units*, which are each marked by a specific suggestion, compliment, or criticism. These categories were not chosen from the outset, but rather are patterns which emerged after an initial review of the data set. We define suggestions as specific measures which the student can take to improve their code, compliments as specific aspects of the student's code which the commenter found novel, elegant, or otherwise impressive, and criticisms as problems with the student's code which lacked specific measures to fix them. Suggestions, compliments, and criticisms are specific, meaning that generic sentiments such as "nice job," "looks good," or "it works" do not constitute a feedback unit. For each comment in the data set, we counted the number of feedback units in each category. In our data set, we distinguish between top-level comments and replies. *Top-level comments* are comments providing feedback to a student's submission, and *replies* are comments which reply to a top-level comment or another reply.

After each of the classes, students were asked to complete an optional survey asking them questions about their experience in the class with peer assessment. The purpose of this survey was to collect data about how helpful students found peer learning to be and to better understand which specific practices they found most useful. We use this data in combination with our analysis of the peer feedback comments to assess the efficacy of our implementation of peer learning. Melville also uses a combination of analysis of peer feedback comments and student survey data to assess the efficacy of their methods for peer learning [7].

Results

Between the eight classes in the data set, there were a total of 1154 top-level comments. Within these comments, we identified a total of 762 feedback units. Of these feedback units, 287 (37.7%) were criticisms, 266 (34.9%) were suggestions, and 209 (27.4%) were compliments.

We discovered that most comments contained few units of meaningful feedback. Of the 1154 top-level comments, 566 (49.0%) contained zero units of feedback, and 464 (40.2%) contained only one. Only 124 (10.7%) of comments contained two or more units of feedback. The mean number of feedback units per comment was 0.65 and the standard deviation was 0.82. A comment containing zero units of feedback did not typically mean that it was blank, but rather that it contained only a generic sentiment which did not constitute a specific suggestion, compliment, or criticism.

Given that threaded comments are a feature provided by Canvas LMS, we wanted to determine whether students make use of the feature, as prior research in the space indicated they might. Knowing whether students make use of this feature could inform decisions about what software institutions may want to use for peer assessment. Of the 1198 comments, only 44 (3.7%) were replies. We attribute this to students only being required to submit a single top-level comment per assignment, which is supported by our finding that students only left an average of 1.1 comments per assignment.

We considered whether there may be a correlation between different categories of feedback, but we found no significant correlations. We did, however, find a correlation between

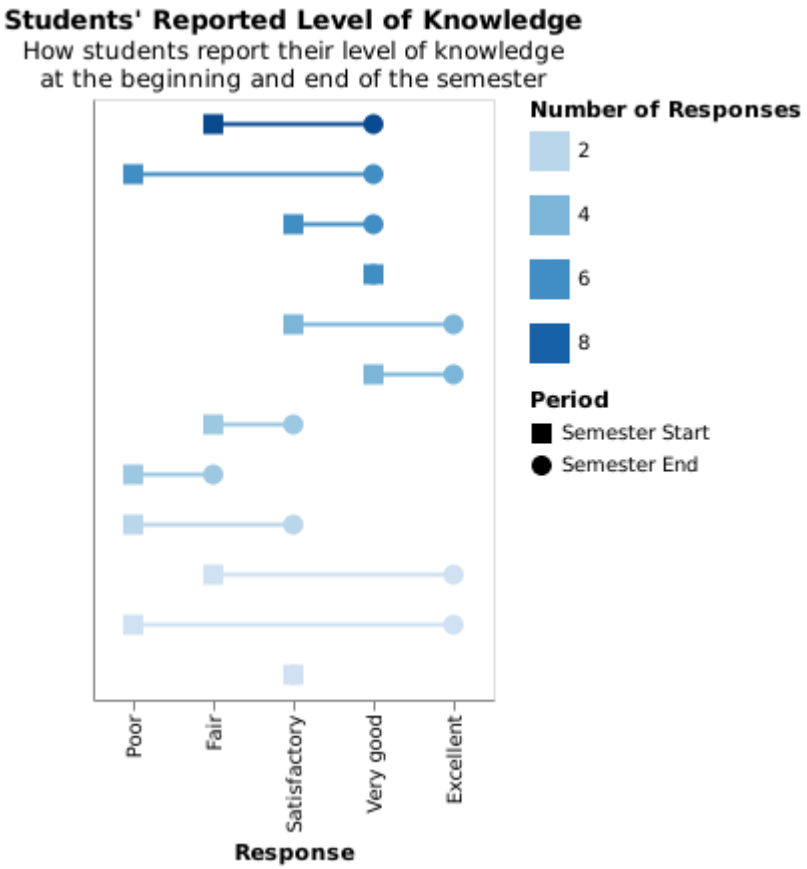
the number of words in a comment and the number of feedback units it contained ($r = 0.69$), indicating, unsurprisingly, that longer comments tended to contain more meaningful feedback.

Given that students were not given the opportunity to resubmit their assignments after receiving peer assessment, we were not able to compare their grades on the assignment before and after receiving feedback. We did, however, find a slight positive correlation between the total number of feedback units provided by a student in a course and their final grade in the course ($r = 0.32$). The direction of causality is unknown. Interestingly, we found that the correlation between the number of feedback units *received* by a student and their final grade in a course was significantly smaller ($r = 0.17$). This could indicate that receiving feedback was less helpful than providing it. It could also be the case that academically stronger students tended to submit comments with more detailed feedback.

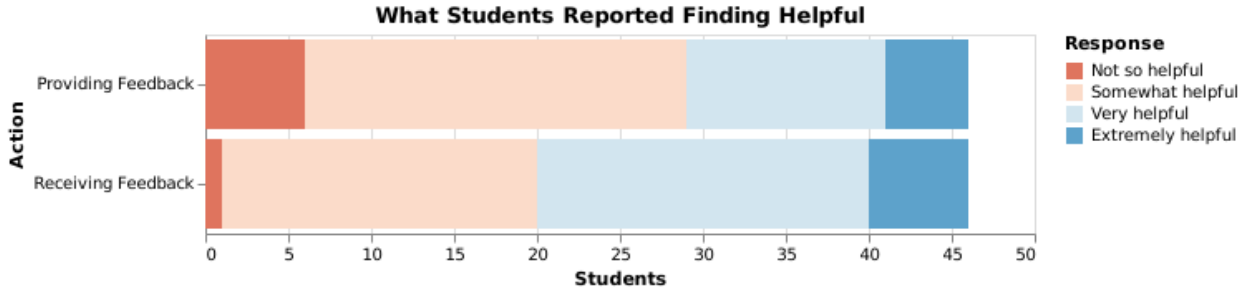
We wanted to determine whether the average number of feedback units per comment changes between courses. Do students tend to leave a lot of feedback in some courses and not in others? Could students leaving detailed comments in a course encourage their fellow students to do the same? A one-way ANOVA test comparing the number of feedback units per comment in each of the 8 classes was performed and we concluded that there was a statistically significant difference between them ($F = 43.0$, $p < .001$). We also wanted to determine if the average number of feedback units per comment changes between assignments, and we also found a statistically significant difference when grouping by each of the 51 assignments ($F = 11.2$, $p < .001$).

We were interested to know whether there is any correlation between the amount of feedback received by a student in a class and the amount of feedback they provide to other students. We found a moderate positive correlation between the number of feedback units received by a student in a class and the number of feedback units provided by a student in a class ($r = 0.57$).

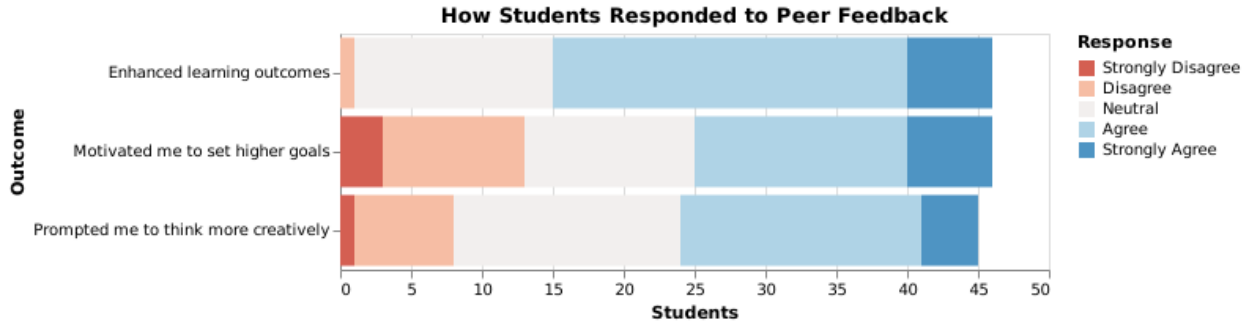
In an optional survey sent to all students upon their completion of the course, which was completed by a total of 46 of them, we asked students to rate their level of knowledge on the course's subject matter at the beginning and end of the course on a five-point likert scale. Nearly all students reported an improvement in their level of knowledge, with the most common response being "Fair" at the start of the course and "Very good" at the end. The distribution of different responses is visualized in the graph below.



In the survey, we also asked students to rate on a four-point likert scale how helpful they found providing feedback to be and how helpful they found receiving feedback to be. Generally, students reported receiving feedback to be more helpful, and the results are visualized in the graph below.



We also asked students to rate their experience with peer assessment in multiple categories on a five-point likert scale, and their responses are visualized in the graph below. 67.4% of respondents agreed or strongly agreed that peer feedback enhanced learning outcomes, 46.7% agreed or strongly agreed that peer feedback motivated them to set higher goals, and 45.7% agreed or strongly agreed that peer feedback prompted them to think more creatively.



Discussion

The nature of the classes involved in this study prohibit some of the more iterative and collaborative forms of peer assessment described in other research such as van den Berg et al., Xu et al., and Melville. The peer learning system implemented for this study was focused more on exposing students to a variety of alternative approaches to problems and less on peer assessment and collaborative learning. We found that in an introductory programming class that involves straightforward programming assignments, there is less opportunity for many distinct solutions to the problem and less opportunity for iterative improvement of those solutions. A concern we had in designing these classes was that taking a more collaborative approach to peer learning would have a homogenizing effect on the students' submissions, reducing the diversity of solutions proposed by students. Melville identifies this type of "groupthink" as a potential challenge of implementing peer assessment in the classroom [7]. One of the goals of implementing peer learning in these classes was to provide students with a diverse range of alternative solutions to the problem to learn from, and it was determined that implementing a more collaborative approach to peer learning did not serve this goal. Additionally, the assignments in this class were graded using a pass/fail system, which prohibits approaches to peer learning that rely on incremental improvement of student submissions as students who have already received a passing grade for the assignment have no reason to resubmit it.

The primary goal of this study was to analyze the efficacy of implementing a system of peer learning that doesn't require substantial restructuring of an existing course and is suitable for disciplines beyond writing. It is because of these requirements that taking a more collaborative approach to peer learning was not feasible for this study, so it's worth analyzing how this constraint influenced our results. We believe the lack of collaboration in our implementation of peer learning explains the general lack of back-and-forth discussion between students that we observed in the results. Educators aiming for a high level of student engagement in peer assessment in the classroom should take note that a more collaborative approach to peer learning may be required to achieve this goal. In particular, assignments which are not designed around collaborative editing or incremental improvement in response to feedback may not be conducive to student engagement in peer assessment. We recommend educators choose assignments which have room for many distinct solutions to the assigned problem, provide opportunities for incremental improvement, and are graded in a way in which students are incentivized to improve and resubmit their solutions.

Because the programming problems assigned to students in this study did not lend themselves to collaborative editing or iterative design, students did not have the opportunity to incorporate peer feedback into their work. Given that students did not have the opportunity to revise their work based on peer assessment, they had little reason to engage with their peers and ask clarifying questions. Educators looking to maximize the benefits of peer learning may wish to design assignments which are well-suited to iterative design and collaborative editing rather than try to adapt existing assignments to make use of peer learning.

The low number of feedback units per feedback comment and the lack of back-and-forth discussion observed in this study could be explained by this study's focus on simple peer assessment and learning from other students' submissions rather than the more iterative and collaborative approaches employed by other educators. While other studies, such as Melville, have focused on how the tools used for peer assessment may affect outcomes, the tools used in this study do not seem to be the limiting factor; Canvas LMS provides a feature to facilitate threaded discussions, but students did not make use of it [7]. While promoting detailed feedback and back-and-forth discussion was not a primary goal of the peer learning system implemented in this study, educators for whom this is a goal may want to consider utilizing one of the more iterative and collaborative approaches to peer learning described in other literature.

Our results found significant variance in the amount of feedback provided by students between classes and even between assignments. Given that the classroom policies used were consistent between classes taught as part of this study, we hypothesize that this variance could be attributed to a "culture of collaboration" in which students are encouraged to leave detailed, meaningful feedback when they see their peers do the same. Conversely, it could be the case that seeing brief, insubstantial feedback from their peers has a discouraging effect and leaves students little incentive to put any meaningful effort into their own feedback. The positive correlation we found between the amount of feedback received by a student and the amount of feedback provided could indicate that students are particularly motivated to leave more detailed feedback when they themselves receive detailed feedback. These findings may have significant implications for educators looking to utilize peer learning in their classrooms. Taking measures early on to actively encourage students to leave detailed feedback for their peers could create a positive feedback loop of encouragement, prompting students to leave more feedback overall.

One interesting finding of the survey students were asked to complete is that students on average reported receiving feedback to be more helpful than providing feedback. However, the amount of feedback provided had a significantly higher correlation with grades than the amount of feedback received. This inconsistency could indicate that providing feedback does not in fact improve grades and that an alternative explanation—like that academically strong students tend to leave more detailed feedback or that academically weak students with less experience have less feedback to provide—is more probable. It could also be the case that there is a disconnect between what students believe to be helpful and what actually has a positive impact on their academic performance, which may be an area for future study. A limitation of the methodology employed for finding this correlation is that we only took into account the final grade for the course. Given that students were not given the opportunity to resubmit their assignments after receiving feedback, it is difficult to track how feedback may have influenced grades on individual assignments.

Because one of the goals of the peer learning system implemented in this study was to minimize the additional workload put on students and instructors, students were not provided with any specific guidance or prompting for providing feedback. Students were not instructed on how to provide feedback or what their feedback should look like and were instead given the freedom to leave free-form responses for their peers. This policy could explain the general lack of detail in the received feedback, indicating that a more structured approach is worth considering for educators looking to maximize the benefits of peer learning.

There is room for future research in this area, and to that end we provide several suggestions for ways the methodology of this study could be adapted for future research. While much of the peer assessment provided by students was subjective, it could be worth classifying feedback according to factual accuracy or relevance. We made no attempt to make such assessments in our research, but it's reasonable to assume that undergraduate students in introductory classes could provide inaccurate feedback, and it may be worth investigating how this could influence outcomes. One concern that faculty sometimes voice when talking about incorporating peer assessment into their classes is that students may provide incorrect or misleading feedback, and future research could investigate how much credence these claims have. One potential pitfall with such an approach is that it requires making judgement calls on the factual accuracy and relevance of students' subjective feedback. Even among experienced software engineers, judgements on the quality of code can often be a matter of personal preference and subjective opinion.

Additionally, there could be room for a more robust system of feedback classification that takes into account the function and subject of the feedback, much like the system used by van den Berg et al. [5]. In our research, we found existing systems for classifying peer assessment, however they were largely designed for collaborative writing courses, and we determined that such a system would not translate well to feedback on computer code. Future research could attempt to design a classification system specifically for this purpose. Because promoting long-form, detailed feedback was not a goal of the peer learning system implemented for this study, and the feedback students provided was generally fairly brief, implementing a more robust system of classification was not necessary. However, future research which focuses more on detailed peer assessment could benefit from a more robust system of classification. In the software engineering field, formal code reviews are common practice among professional developers, and future research could analyze the methods used in these sorts of code reviews to develop a system for classifying feedback.

Conclusion

The main objective of the peer learning system implemented in this study was to expose students to a variety of alternative solutions to the problems they were assigned so they could learn from their peers through example. The peer learning system in this study required minimal changes to the course curriculum and didn't impose significant additional responsibilities on the students or educators. Additionally, the courses in this study were not designed as collaborative or feedback-centric courses, but rather more traditional courses that incorporated elements of peer learning.

That being said, we see room for improvement in the level of student engagement and the quality and quantity of peer assessment observed in this study. Prior research has found that certain styles of peer learning lend themselves to better student engagement, however there are certain challenges which must be overcome in order to apply these techniques to an introductory programming course. Based on our experience, we make the following suggestions to educators:

1. Design a peer assessment system that encourages and incentivizes students to engage with their peers in back-and-forth discussion. Encourage students to reply to their peers' comments and ask clarifying questions, or provide incentives for them to do so.
2. Make use of structured feedback. Educate students on how to provide meaningful and substantive feedback, and provide them with templates to provide structured feedback to their peers.
3. Design assignments which promote collaboration. Give students the opportunity to incorporate their peers' feedback into their submissions or work collaboratively with their peers.
4. Create a positive culture of collaboration. Enact policies to promote detailed, meaningful feedback early on in order to create a positive feedback loop of encouragement.

Sources

[1] K. J. Topping, "Trends in Peer Learning," *Educ. Psychol.*, vol. 25, no. 6, pp. 631–645, Dec. 2005, doi: 10.1080/01443410500345172.

[2] H. M. Levin, G. V. Glass, and G. R. Meister, "Cost-Effectiveness of Computer-Assisted Instruction," *Eval. Rev.*, vol. 11, no. 1, pp. 50–72, Feb. 1987, doi: 10.1177/0193841X8701100103.

[3] K. Topping and S. Ehly, *Peer-assisted Learning*. Routledge, 1998.

[4] K. Topping, "Peer Assessment Between Students in Colleges and Universities," *Rev. Educ. Res.*, vol. 68, no. 3, pp. 249–276, Sep. 1998, doi: 10.3102/00346543068003249.

[5] I. van den Berg, W. Admiraal, and A. Pilot, "Designing student peer assessment in higher education: analysis of written and oral peer feedback," *Teach. High. Educ.*, vol. 11, no. 2, pp. 135–147, Apr. 2006, doi: 10.1080/13562510500527685.

[6] A. Xu, H. Rao, S. P. Dow, and B. P. Bailey, "A Classroom Study of Using Crowd Feedback in the Iterative Design Process," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA, Feb. 2015, pp. 1637–1648, doi: 10.1145/2675133.2675140.

[7] N. P. Melville, "Crowd-Sourced Peer Feedback (CPF) for Learning Community Engagement: Results and Reflections from a Pilot Study," in *2014 47th Hawaii International Conference on System Sciences*, Waikoloa, HI, Jan. 2014, pp. 32–41, doi: 10.1109/HICSS.2014.14.

[8] J. van der Pol, B. A. M. van den Berg, W. F. Admiraal, and P. R. J. Simons, "The nature, reception, and use of online peer feedback in higher education," *Comput. Educ.*, vol. 51, no. 4, pp. 1804–1817, Dec. 2008, doi: 10.1016/j.compedu.2008.06.001.

[9] W. M. Saunders, "Collaborative writing tasks and peer interaction," *Int. J. Educ. Res.*, vol. 13, no. 1, pp. 101–112, Jan. 1989, doi: 10.1016/0883-0355(89)90019-0.

[10] D. J. Nicol and D. Macfarlane-Dick, “Formative assessment and self-regulated learning: a model and seven principles of good feedback practice,” *Stud. High. Educ.*, vol. 31, no. 2, pp. 199–218, Apr. 2006, doi: 10.1080/03075070600572090.
[11] M. Shekary and M. H. Tahririan, “Negotiation of Meaning and Noticing in Text-Based Online Chat,” *Mod. Lang. J.*, vol. 90, no. 4, pp. 557–573, 2006, doi: https://doi.org/10.1111/j.1540-4781.2006.00504.x.