

Introduction to Student Peer Review

Prepared by: Jessica Dawson, Ian Mitchell

[Computer Science Department](#) and [CS-CWSEI](#), University of British Columbia

November 18, 2014

Overview

This document has been prepared for the Department of Computer Science at the University of British Columbia. It presents a brief introduction to the concept of student peer review assignments and the different types of peer review, as well as the advantages and disadvantages reported in the literature and by members of the department who have used peer review in their own classes. The information was compiled through a broad literature review and conversations with instructors in the department who have had experiences implementing peer review in their courses.

A companion document, [Resources and Guidance for Student Peer Review](#), provides some guidance to instructors who want to implement peer review in their own course, and includes a detailed list of software systems for peer review, as well as references to resources where readers can look for more detailed information.

What is peer review?

Student peer review is modeled after scholarly peer review. Peer reviews can be used for a variety of assignments and activities, including essays, reports, programming assignments and code reviews, and can be performed instead of *or* in addition to review by course staff.

In most student peer review processes students submit their own work first and then evaluate the work submitted by one or more of their peers. Generally the assignments are randomly assigned to reviewers, and all assignments and reviews remain anonymous (double blinded review).

The process can be adapted to include many additional stages such as training, calibration, revision, self-assessment, or feedback on the reviews themselves. The process can also be used for multi-stage assignments, where the reviews are completed ‘in-flow’ between each stage and students have the opportunity to integrate feedback on initial step(s) before completing the subsequent step(s) [1].

Peer review has been used by a small number of instructors in the Computer Science department for courses that include CPSC 110, CPSC 430, CPSC 540, SCIE 113.

Types of Student Peer Review

Many variations of peer review have been proposed, and the best type to use will vary with the learning goals of the assignment and constraints on resources. The terms *peer evaluation* and *peer review* are often used interchangeably to describe the general practice of students reviewing each other’s work; however, there are two distinct ways in which the student evaluations / reviews can be used:

Peer assessment (also frequently called *peer grading*): Assessment from peers is used either alone or in combination with course staff grades to determine a final grade for the assignment. Generally requires (at a minimum) assigning numerical ratings based on detailed criteria and rubrics.

Peer feedback: Peers review the quality of the work and provide written and/or numerical feedback, but it does not contribute to assignment grades.

In all cases, the nature of the evaluation can vary greatly. In peer feedback processes, reviewers will typically provide comments and discussion on the quality of the work, either generally, or when measured against given criteria. Peer feedback tends to be more qualitative in nature, but can be very precise and include numerical scoring. In peer assessment, the evaluations are necessarily quantitative in order to support calculation of marks, and often require more detailed marking criteria and rubrics. However, peer assessment can also include qualitative comments and discussion. A range of arguments for and against both styles, as well as recommendations for how they should be combined, can be found throughout the literature (e.g., [2]).

Advantages to Student Peer Review

This section outlines many advantages of peer review discussed in the literature, and in particular focuses on those highlighted by members of the department. Most courses that use peer review in computer science use it as one method to determine grades, but many of the following advantages also accrue when peer review is used solely to obtain feedback.

Improve Student Performance

Peer review activities have been shown to improve writing performance in many disciplines [3]–[5], but have also been shown to improve performance on other types of activities, including programming assignments [6] and creation of test suites [7]. Numerous studies have shown larger improvements in student learning through the peer review process than through feedback from instructor or subject matter experts alone [8]. For students who start off with poorer performance, peer review has also been shown to have a much larger impact on their improvement over time [9]. These outcomes are not surprising, as peer review tends to expose students to a range of examples of quality, especially when combined with effective calibration exercises that define boundaries and expectations around quality.

Expose students to other perspectives and expand learning opportunities

Students also benefit from the act of creating and giving reviews [10]. The activity forces students to see the work they produce and the task that they've been given from the perspective of an evaluator. It also exposes students to a broader selection of perspectives, solutions, strategies and insights than they would see otherwise, even if they also receive expert feedback from course staff. For example, in CPSC 540, peer assessment has been adopted in part as a way to expose students to other ways of solving machine-learning problems that they may not have considered in their own solution.

Define and clarify learning outcomes

To effectively perform peer review, students need to develop a reasonable understanding of what they are trying to evaluate and the parameters on which they can and should make judgments. To effectively support peer assessment, for example, instructors must be able to explain how they arrive at a mark. Peer review therefore has the benefit of encouraging professors to clearly define the learning outcomes for an assignment and the criteria that they expect to be met. Engaging in this process with students can lead to a greater shared understanding of quality, upon which students can rely when deciding what they need to do to achieve a particular outcome on their own submission [2], [11].

Further, evaluating the work of others according to this shared understanding reinforces why the reasoning behind why these particular criteria or expectations are important. Instructors of CPSC 110 cited this as a particular benefit to using peer assessment for marking coding style, an issue that students tend to only recognize the importance of once they have tried to interpret and understand someone else's code.

Reduce marking workload on course staff

Peer assessment can also reduce the time the course staff spends on marking [11]. For example, using peer assessments to determine final grades for assignments and limiting the efforts required by the instructors and TAs to spot-checking and reviewing disputes can free up staff time for other activities. The use of peer assessment may also allow instructors to offer more assignments than would be practical for course staff to grade directly, which can provide more opportunities for practice and iterative improvement of skills. In CPSC 430, peer assessment has allowed the instructor to increase the number of essay assignments from two per term to more than ten, thereby increasing the number of opportunities for students to receive feedback and improve their essay writing skills.

Expand amount and speed of feedback

Given the pressures of expanding enrollment in the department, peer review can be especially useful in large classes when opportunities for instructors and TAs to give individualized feedback are highly restricted by time and resources. In CPSC 110, some instructors commented that the use of peer assessments with detailed rubrics supports more detailed and specific feedback than would be possible with TA only marking. Further, because students review only a handful of assignments, feedback can often be turned around much more quickly than when course staff must mark all assignments. It can also therefore be reasonable to require peers to provide more detailed qualitative feedback (such as justification for assigning a particular grade) than there is time for course staff to provide.

Challenges to Student Peer Review

Although successfully implementing peer review in a course can reduce time and effort needed by instructors in some areas, it can still require significant time and effort upfront and throughout the term. There are also a number of challenges associated with peer review that instructors may encounter and that should be considered when deciding how to implement peer review in a course.

Previous reviewing ability and experience

Student peer reviewers are typically novices in their disciplines. Furthermore, many students will have little experience in effectively and helpfully reviewing or critiquing the work of their peers, and thus the feedback and evaluations they generate are more likely to be inaccurate than feedback from trained instructors or TAs. Extending peer review to include peer assessment poses additional challenges, including maintaining validity and reliability in the grading and minimizing the influence of “rogue reviewers”[6], [12]. However, numerous studies have shown that with training and careful design of the peer review process, students are capable of assigning grades that are internally reliable, and highly valid with respect to staff-assigned grades [4], [5], [7].

To be successful, students must be trained in how to provide high quality reviews. Instructors need to provide detailed and precise instructions on expectations for both the assignments and the reviews. In the case of peer assessment, such instructions typically take the form of detailed point-by-point rubrics that leave little ambiguity or room for interpretation [4]. To be most effective, this guidance should be

combined with opportunities for the students to practice their reviewing skills through training exercises and/or increasing the overall number of assignments.

One common training strategy is “calibration,” wherein students review assignments that have been previously assigned a consensus grade by course staff; students can then be given feedback on how close their assessment came to the *true* grade. Some software systems also use such calibrations as a measure of review quality in assigning grades or deciding how students will proceed in the review process. It is similarly important to put in place mechanisms for tracking review quality; for example, having course staff mark reviews, allowing student appeals, and/or performing random spot checks.

In CPSC 430, calibration activities are used to train students on how to grade their peers’ assignments effectively and accurately. In the course, both reviews and writing assignments have grades assigned. Students are separated into one of two pools, “independent” or “supervised,” in one of two ways: to be “independent” they must either receive sufficiently high marks on their previous week’s reviews, or they must practice on calibration essays until they demonstrate consistent reviewing performance.

Assignments and reviews by students in the supervised pool are marked by course staff; assignments in the independent pool are peer graded, and students in this independent pool automatically receive full marks for their reviews. This approach has been effective in training students to create high quality and consistently accurate reviews, as students are motivated to get into the independent pool as soon as possible. To maintain review quality, a random selection of reviews and all reviews that assign very high scores are spot checked by course staff.

Negative student perceptions of peer review

Research has frequently reported that a negative perception of peer review, and peer assessment in particular, is fairly common among students [2], [5], [13]. Students tend to doubt the ability of their peers as compared to their TAs and instructors [13], and may question whether or not the feedback is valid. Students also tend to perceive peer review as being less fair than review by subject matter experts.

The same strategies that can be used to improve students’ ability to peer review – such as clear and precise rubrics, calibration and training, and increasing the number of activities and peer review experiences – can also improve their perceptions of the process.

Some software systems attempt to address fairness by weighting grades given by a reviewer more or less depending on the accuracy of their calibrations or final reviews (e.g., [12]). Anecdotally, instructors of SCIE 113 have reported that while weighted grades can contribute to positive perceptions of fairness for some students, they can actually enhance negative perceptions when the system is complex and not transparent. Another strategy for increasing positive perceptions is to involve course staff in marking. For example, SCIE 113 involves students in peer assessment of early drafts of assignments, but leaves the marking of the final paper to the instructor. Multiple instructors have also commented on the importance of having an appeals process in place for peer graded work; even if students do not regularly make appeals, having the option to do so seems to increase the perceived fairness of the process.

Student motivation to complete peer reviews

Students may not be particularly motivated to complete peer reviews, especially if they have a negative perception of the process. Repeated use of peer review can address this problem, as can giving students opportunities to revise their work to incorporate the feedback they receive. Students may begin to find the activity more valuable over time, especially if they see improvement in their own work and grades as a result.

Including a grade for completing reviews is, not surprisingly, also an effective way to motivate students to engage in the review process. However, just giving students grades for completing the reviews does not ensure that they will be of decent quality. Having course staff mark reviews is probably the most effective way to ensure quality when using this approach, but adds considerable time and effort. An alternative approach is software systems that use weights based on factors such as performance and accuracy on calibration reviews, or similarity to other reviews of the same assignment to assign grades for review quality. Some software systems also allow students to give feedback on the helpfulness of the reviews that they received, which could then be factored into a grade for the review.

Student conduct and professionalism

Instructors I spoke to noted that, although rare, some reviews include rude or inappropriate comments. Thus, there must be mechanisms planned so that students can report such comments. In addition, students who have limited experience giving feedback and critiques may not realize that they are being rude. It can therefore be helpful to include a discussion of professionalism and examples of appropriate and inappropriate comments when training students in the peer review process.

Plagiarism

When using peer review for assignments in which there will be repeated revisions or multiple stages, plagiarism is a concern for some instructors. However, as multiple instructors in the department noted, having students see other peoples' code is representative of real life and can be beneficial (see also [1]). One option is to weight grades for earlier submissions more than the later revised work—students still get the benefit of peer review, but are motivated to do as good a job as they can in the first round.

Straight plagiarism can be detected in final drafts by having course staff review or mark the submissions in addition to peer review, or by employing plagiarism detection tools like TurnItIn. However, as noted by instructors in SCIE 113, it can be much more difficult to detect instances where students have copied other students' good ideas, and this concern can lead to negative opinions of peer review by students.

Time and effort required by instructor and students

The peer review process takes time for both students and course staff to master. Approaches that are meant to improve student attitudes to peer review and the quality of their reviews can be successful, but take nontrivial time and effort. For this reason, most instructors felt that peer review was not really worth the effort for only one or two assignments, regardless of whether or not they felt that the peer review activities they tried in their courses were successful.

Summary

Student peer review brings many advantages to the learning environment, and can be adapted to suit a range of learning goals and types of assignments. Although the process has challenges, most can be reduced or solved through the use of software tools and by tailoring the peer review process to support the specific goals the instructor aims to achieve. In particular, all instructors interviewed agreed that in order for students to create quality peer reviews, and in particular peer assessments, they must receive guidance in how to evaluate through training, calibration, and detailed marking schemes and rubrics.

For more information on software tools see the companion document:

www.cwsei.ubc.ca/resources/files/Student-Peer-Review_Resources.pdf

Acknowledgements

Many thanks to Ed Knorr, Meghan Allen, James Wright, Kevin Leyton-Brown, Ron Garcia and Mark Schmidt for sharing their experiences with peer review in their courses.

Annotated Bibliography

- [1] J. G. Politz, D. Patterson, S. Krishnamurthi, and K. Fisler, “CaptainTeach: Multi-Stage, In-Flow Peer Review for Programming Assignments,” in *To Appear in Proc. ITiSCE '14*, 2014.

Companion paper to [7]. Presents a method and accompanying programming environments that supports peer review of deliverables at multiples stages within in-progress assignments. The authors argue for the benefits of supporting feedback throughout the process of designing a software program (such as on specifications, test suites and code) and report on their experience using this method in two undergraduate computer science courses.

- [2] N.-F. Liu and D. Carless, “Peer feedback: the learning element of peer assessment,” *Teach. High. Educ.*, vol. 11, no. 3, pp. 279–290, 2006.

The authors argue that evaluations of peer review have overly focused on the use of peer assessment, leading to invalid conclusions about the overall value of peer review activities and resistance to their adoption. They discuss a range of benefits of peer feedback, and argue that it is a worthwhile activity in and of itself, and as an effective precursor to successfully introducing peer assessments.

- [3] R. D. Gerdeman, A. A. Russell, and K. J. Worden, “Web-Based Student Writing and Reviewing in a Large Biology Lecture Course.,” *J. Coll. Sci. Teach.*, vol. 36, no. 5, pp. 46–52, Nov. 2006.

A study of the use of the Calibrated Peer Review software system to support peer assessment on writing assignments in three large biology classrooms. Results show an increase in writing performance and improvement in reviewing skills (measure as deviation in assigned scores) over the course of three assignments.

- [4] C. Kulkarni, K. P. Wei, H. Le, D. Chia, K. Papadopoulos, J. Cheng, D. Koller, and S. R. Klemmer, “Peer and self assessment in massive online classes,” *ACM Trans. Comput. Interact.*, vol. 20, no. 6, pp. 1–31, Dec. 2013.

Reports on the authors’ experiences designing and deploying peer assessment over two iterations of an online introduction to Human Interaction Class in Coursera, which was the first large-scale course to do so. The authors present the results of three experiments to improve grading accuracy, and discuss approaches that led to decreased variance in student scores.

- [5] C. Kwangsu, C. D. Schunn, and R. W. Wilson, “Validity and Reliability of Scaffolded Peer Assessment of Writing From Instructor and Student Perspectives,” *J. Educ. Psychol.*, vol. 98, no. 4, pp. 891–901, 2006.

A study of the validity and reliability of peer-assessment of writing assignments by 708 students across 16 different courses from 4 universities using the SWORD system. Results show that the aggregate ratings of four or more peers were highly reliable in terms of similarity within one another, and were also highly valid when compared with instructor ratings of the same papers. Despite these findings, the authors report very poor perception of the reliability and validity of peer review among students.

- [6] H. Sondergaard, “Learning from and with peers: the different roles of Student Peer Reviewing,” in *ACM SIGCSE Bulletin - ITiCSE '09*, 2009, vol. 41, no. 3, p. 31.

Reports on the authors pedagogical philosophy on peer reviewing, and on his experience using peer review in a Computer Science course on compilers.

- [7] J. G. Politz, S. Krishnamurthi, and K. Fisler, “In-Flow Peer-Review of Tests in Test-First Programming,” in *To Appear in Proc. ITiSCE '14*, 2014.

Companion paper to [1]. The authors present the results of an experiment on the peer review of test suites of in-progress assignments. The authors find evidence that the process led to improvements in the final test suites submitted by students, and find that the peer-assigned grades were highly consistent with staff-assigned grades.

- [8] J. A. Rudd, V. Z. Wang, C. Cervato, and R. W. Ridky, “Calibrated Peer Review Assignments for the Earth Sciences,” *J. Geosci. Educ.*, vol. 57, no. 5, pp. 328–334, 2009.

A study conducted in a large introductory science course. As part of the study, students in one section of the course were given a CPR assignment, while students in another section wrote an essay that was instructor graded with extensive feedback. Results showed that students who completed the CPR assignment showed more improvement over the students whose essay was marked by the instructor.

- [9] A. B. Gunersel, N. J. Simpson, K. J. Aufderheide, and L. Wang, “Effectiveness of Calibrated Peer Review for improving writing and critical thinking skills in biology undergraduate students,” *J. Scholarsh. Teach. Learn.*, vol. 8, no. 2, pp. 25–37, 2008.

A study that examines improvement in scores and reviewing ability over repeated uses of Calibrated Peer Review for writing assignments in a Biology course, where very limited additional feedback was provided by the instructor. Lower performing students showed more improvement in writing skills and reviewing skills over time than students who were higher performing, and the difference between the two groups decreased over the course of the study.

- [10] K. Lundstrom and W. Baker, “To give is better than to receive: The benefits of peer review to the reviewer’s own writing,” *J. Second Lang. Writ.*, 2009.

This study investigates which aspect of peer reviewing – giving feedback or receiving feedback – does more to improve writing performance. Students were divided into two groups over the semester, one which gave reviews and one which received reviews; results showed that students

who focused on giving feedback over the course of the semester made larger gains in writing improvement than those who only received feedback.

- [11] G. Gibbs, "Using assessment strategically to change the way students learn.," in *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, S. Brown and A. Glasner, Eds. 1999.

A discussion of how different assessment methods can be used strategically to change the ways in which students learn, which includes examples of how peer assessment has been used in real classes. The author presents peer assessment as a method for reducing marking workload and increasing the number of assignments and total amount of feedback that can be given in a large class, as well as helping students to internalize expectations of quality criteria and apply them to their own work and the work of their peers.

- [12] J. Hamer, K. T. K. Ma, and H. H. F. Kwong, "A Method of Automatic Grade Calibration in Peer Assessment," in *Proc. ACE '05*, 2005, pp. 67–72.

The authors investigate concerns of plagiarism and the influence of rogue reviewers on the validity and reliability of peer assessment. They present an algorithmic method for identifying and minimizing the impact of outlier reviews on results, as well as simulations suggesting that this method is effective.

- [13] J. H. Kaufman and C. D. Schunn, "Students' perceptions about peer assessment for writing: their origin and impact on revision work," *Instr. Sci.*, vol. 3, pp. 387–406, 2011.

A large scale study of student perceptions of peer reviewing and an attempt to identify factors that lead to negative perceptions of the process. The authors find that negative perceptions are influenced by a belief that other students are not qualified, or that the process is unfair. They also find that students' opinions of the process improve after performing peer assessments and, not surprisingly, that perceptions shift positively to a greater extent when the feedback they receive is positive and useful.