Calibrated Peer Review

Scenario

In a workshop over the summer, Danielle worked with actors, writers, scene designers, costume designers, and makeup artists to create a fledgling company dedicated to making short videos. This semester, she is taking a class called “Movie Making in Miniature.” Because the course uses Calibrated Peer Review, Danielle hopes it can provide an unbiased audience to contribute anonymous feedback.

For her first assignment, Danielle submits a “10-minute mystery” that she produced. She uploads her file and turns to the Calibrated Peer Review tutorial, which explains how to review videos by giving a score for each question in a provided rubric. Danielle is then ready for the calibration process. For this, she watches a sample video that explains how to plant an orange tree. The rubric was built by the instructor, who provided his responses for comparison. She reviews the orange-tree video, a documentary, and a video lecture with animation. She’s delighted to find that the scores she gave these works are close to those of the instructor, which earns her a high rating as a reviewer.

Using the same rubric she used on the samples, Danielle rates the work of her peers, reviewing five movies, often adding comments to explain her rating. When her reviews are finished, she uses the same criteria to judge her own work. She now has a clearer sense of the strengths and weaknesses of her video: a continuity error seems obvious, and the audio is uneven, but her movie is more polished than most.

At the end of the week, she receives peer feedback and signs on to Facebook to discuss the results with her creative team. The ratings are high for execution but lower for “raised awareness of an issue or informed the viewer.” The team found the comments made by fellow movie makers to be extremely helpful, some of whom said the mystery was a hit. Her production group will recut the video to address some of the points in question and make a change to a piece they are just completing. As she looks over the next video she will be submitting, Danielle finds she has already sharpened her critical eye toward her own work.

1 What is it?

Calibrated Peer Review (CPR) is a system for coordinating and evaluating peer reviews of student work. CPR was developed at UCLA—based on general principles of peer review—to allow instructors in the sciences to incorporate more writing assignments into their classes without spending additional time in assessment and markup. CPR has evolved to be discipline-independent, and today students can use it to review not just essays but also such content as works of art, oral presentations, posters, graphs, or computer code. The system provides double-blind anonymity for submissions and critiques, and it offers instructors an online archive of assignments and an authoring tool for creating their own exercises.

2 How does it work?

In a course that uses CPR, students initially take a tutorial on how to review peers’ work. Students then evaluate two or three sample assignments, using a rubric set up by the instructor. The rubric consists of a series of questions about the work. Each student’s ratings are then compared to those assigned by the instructor (which constitute the “correct” ratings), and the proximity of the student’s evaluation with the instructor’s determines the student’s score as a reviewer. Once students’ reviews are thus “calibrated,” students begin assessing each other’s work.

The CPR system randomly distributes student-submitted assignments to the other students in the class for evaluation. Reviewers, who cannot see who submitted the assignment, provide ratings per the rubric and can also add comments. (Students also use the rubric to review their own assignments.) The ratings and comments—as well as the calibration scores of the reviewers—are then shared with the student who submitted the assignment. Although students cannot see the names of their peers at any point in the process, all assignments and all reviews, as well as the names of students who completed them, remain visible to the instructor.

3 Who’s doing it?

CPR has been used by hundreds of universities since its launch. Often these uses involve a few classes, as with an introductory course in geology at Carleton College, first-year
science courses at the University of British Columbia, chemistry courses at Texas A&M, or biology courses at the University of Oklahoma. Many such implementations have spawned small-scale studies of CPR’s effectiveness with results that have varied, perhaps influenced by what skills were measured, what work was submitted, and how measuring was done. One such study, for example, involves large introductory political science classes at the University of North Texas, where researchers concluded that CPR could improve students’ writing skills and promote higher-order learning. Recently, the adoption of peer reviews by Coursera has helped to focus interest on this method of providing assessment in MOOCs.

4 Why is it significant?
Peer review involves and engages reviewers in the work of their colleagues, and CPR does this in an anonymous, online, and global system. In MOOCs, peer assessment is one of few viable options for formative assessment and feedback, and it can be used in a similar manner with any community of learners, large or small. Students get to see their work through the eyes of anonymous reviewers. They can compare the evaluation of others with their own assessment, done with the same rubric, and even make an appeal to the instructor for review where appropriate. As an added benefit to those who provide the reviews, research has shown that when students review multiple submissions on the same topic in CPR, they absorb content related to the assignment as part of their effort. Reviewers tend to reflect on what they read and develop the habit of critical thinking, while refining skills such as assessing, analyzing, and framing a logical argument.

5 What are the downsides?
The value of peer review depends on certain assumptions, one of which is that the scores given by the instructor in the calibration exercise are the “right” answers and that a student’s conformity to these responses indicates his or her value as a reviewer. This notion conflicts with the “multiple viewpoints” that many consider the primary value of peer review. Another assumption is that those who are reviewing the work are making an honest effort to give the best evaluation they can. Because the CPR system assigns raters to a work, an assignment could be evaluated by five raters who all have low calibration scores, producing feedback that’s not very useful. Similarly, if all raters for a submission miss the deadline, that student would have no peer feedback on the assignment. Some students have complained about the added workload of being reviewers, which might result in cursory reviews or discourage some students from taking courses that include CPR.

6 Where is it going?
UCLA periodically updates the CPR service. In one upcoming feature, the system will track and coordinate initial student drafts for writing assignments and manage the resubmission of subsequent drafts after initial peer reviews. Putting the ratings from a peer review system into a social media environment could encourage discussion among students about the value of comments and scores provided and the value of giving, accepting, and implementing critiques. Wider use of the CPR system could expand the understanding and value of weighted or otherwise normalized peer reviews, encouraging their use in other contexts. CPR or other peer review systems might become a common element in MOOCs and other large-enrollment learning environments. One such model is the “Credibility Index” currently being pursued at Pennsylvania State University, which in addition to evaluating peer raters’ accuracy evaluates their consistency and “transferability”—their accuracy across different assignments and problem contexts.

7 What are the implications for teaching and learning?
CPR offers quick feedback for students and, by casting students as reviewers, creates a dynamic that can help them learn to recognize strengths and weaknesses in their own efforts. The large, anonymous audience provided by CPR can provide an objective perspective on students’ work and an opportunity to accept the views of others about what they create. For MOOCs, peer review presents a viable option for assessment, making these massive online classes a much more viable delivery model for credit-based courses. CPR can generate conversations about responsible criticism and the value of external review. Study groups can talk about the results received from CPR, comparing notes on where their hands-on reviews might be at variance and determining whether the rubric was an effective mechanism for judging the assignment. Such an approach might open the door to student discussions about the differences between critiques that emerge from CPR and those provided by study teams, helping them to recognize the limitations and appreciate the value of each.