



Calibrated Peer Review: Use of Open-Ended versus Closed-Ended Peer Evaluations

Laurel L. Hester, Cornell University and Briana Timmerman, University of South Carolina

Introduction

Peer-reviewed papers are the gold standard in scientific writing (Fletcher and Fletcher, 1997). Peer-review assignments introduce this authentic scientific activity while also encouraging students to develop self-assessment skills and assume responsibility for their own learning (Siebert and McIntosh, 2001). However, students can be reluctant to critically evaluate other students and may devalue criticism from peers whom they view as non-experts (Henderson and Buising, 2002).

Calibrated Peer Review™ (CPR, <http://cpr.molsci.ucla.edu>) is a web-based instructional program that facilitates anonymous peer review and trains student reviewers as they evaluate 'calibration' essays chosen and rated by the instructor. This process gives students practice reviewing three example assignments and lets students compare their ratings with those provided by the instructor (Russell, 2004). Student performance in this calibration phase is used to calculate a 'reviewer competency index' (RCI) based on the degree of matching between student and instructor reviews. This study compares student response to a CPR assignment using closed-ended review prompts ('close-ended CPR') with response to an assignment using more open-ended prompts ('open-ended CPR', see Table 1).

Table 1: Open-ended versus close-ended assignment design.

Question types for scoring calibrations and reviews	Open-ended assignment	Close-ended assignment
Yes/No	0	28
A/B/C* or None/Some/Many	14	2
Required Text Box Comments	12	1

* A = Excellent job; B = Good job, but needs some improvement; C = Needs much improvement

Methods

Students in a large introductory biology class peer-reviewed lab reports (on *Drosophila* Genetics) using CPR™. At the beginning of the process, students were given a grading rubric that included both open-ended assessment descriptors and close-ended questions for each lab report section (see handout). Half of the students uploaded their papers to the 'close-ended CPR' assignment and the other half to the 'open-ended CPR' assignment. After completing the 'calibration' phase, students reviewed three student papers, viewed peer feedback and revised their lab reports for final submission. Teaching assistants (TAs) graded the final lab reports using the grading rubric initially handed out to the students.

The open-ended and close-ended CPR assignments differed as shown in Table 1. Calibration feedback for the open-ended CPR assignment included example text answers. Students in both groups also assigned a score of 1-10 to all calibration and student papers evaluated.

Teaching assistants who had not been involved in the CPR process graded revised lab reports from students participating in both assignment versions. The Mann-Whitney test was used to compare median grades and RCI scores of the two groups. Time spent on a review was also compared for a sub-sample of students from each group. End of semester course evaluations asked students: "comment on CPR: did it help you understand scientific writing?" Responses were scored as positive, mixed/neutral, or negative. Students were also asked whether they found the calibrations or the peer feedback most useful.

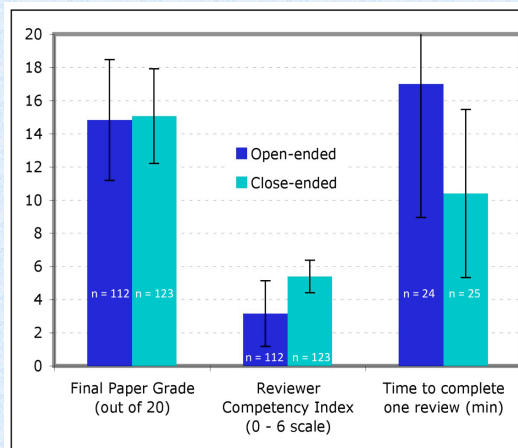


Figure 1: Final paper grades did not differ between students participating in the close-ended versus the open-ended CPR assignments, but reviewer competency index scores and time spent per review were significantly different between these two groups. Means and standard deviations are shown.

- ◆ Final paper grades did not differ between students in the two CPR assignment groups ($p = 0.88$, Figure 1).
- ◆ Students in the open-ended CPR group had a harder time matching instructor ratings on the calibration papers. This led to significantly lower ($p < 0.0001$) Reviewer Competency Index (RCI) scores for students in this group (Figure 1).
- ◆ Students in the open-ended CPR group spent more time on their reviews ($p < 0.005$, Figure 1).
- ◆ Text scores assigned to draft papers in the close-ended CPR group were more highly correlated with final paper grades ($R^2 = 21.2\%$) than for the open-ended group ($R^2 = 9.5\%$, $p < 0.005$ for both regressions).
- ◆ Both student groups had an overall negative evaluation of the process, in many cases due to technical difficulties and the amount of time required. (Figure 2).
- ◆ Overall, students found peer comments more useful than feedback from calibration text scoring. However, significantly more students from the close-ended-evaluation group found the calibration procedure useful (Table 2, Chi-Squared = 7.53, $p = 0.006$).

Table 2: Student responses when asked whether the calibrations or the peer feedback was more useful.

Most Useful Part of CPR process	Open-ended CPR group	Close-ended CPR group
Calibrations	28 (21.0%)	38 (31.9%)
Peer Feedback	109 (79%)	81 (68.1%)
Total	138	119

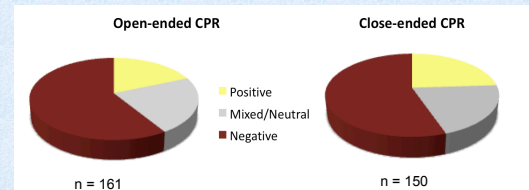


Figure 2: Student response to CPR was mostly negative. Differences between groups were not significant.

Conclusions

Students more reliably rated papers using the close-ended review question format.

Students were better able to match calibration scores with the close-ended question version. The higher correlation between student draft text ratings and final paper grade suggests that student reviewers could also more reliably rate student drafts using the close-ended questions. Thus, although open-ended feedback theoretically allows students to more specifically tailor their comments to the paper being reviewed, close-ended questions may have an advantage over open-ended questions for naïve reviewers.

Frustration with technology, lack of student buy-in and variable peer-ratings led to negative student evaluation of CPR.

This study reveals challenges involved in using technology and in introducing a more student-centered learning environment. Student complaints were typically related to frustration with technology or to lack of student (or TA) buy-in for the procedure: "do you think people actually read all that to do the required reviews?" and "The TAs should review the papers because they are the ones grading them." Of course, some students did appreciate CPR: "Peers were honest and gave specifics on how the paper could be improved. Peer reviewing helped (me) to improve my own paper." For a more complete analysis of student response to CPR in another class, see Waldvoord et al. (2008) who also note student frustration with variable peer grading. Some student complaints critical of peer-reviews can be addressed via clear communication from instructors about the value of peer-review: both as an authentic scientific activity and as a mechanism for improving student writing and editing skills.

Assignment structure significantly impacts CPR experience.

Our results suggest that there may be a trade-off between increased reviewer consistency achieved with close-ended questions and promotion of higher-order analysis skills with open-ended questions. In cases where CPR is being used as a form of peer-grading, reviewer consistency is an especially important consideration. Instructor communication about the goals of student peer-review is also essential because constructionist approaches typically require more effort on the part of learners.

References

- Fletcher, R. H., & Fletcher, S. W. (1997). Evidence for the effectiveness of peer review. *Science and Engineering Ethics*, 3, 35-50.
- Henderson, L., & Buising, C. (2002). A peer-reviewed research assignment for large classes. *Innovative techniques for large-group instruction* (pp. 30-34). Arlington, VA: NSTA press.
- Russell, A. A. (2004). Calibrated peer review - A writing and critical-thinking instructional tool. In M. Druger, E. D. Siebert & L. W. Crow (Eds.), *Teaching tips innovations in undergraduate science instruction* (pp. 54). Arlington, VA: NSTA press.
- Siebert, E. D., & McIntosh, W. J. (Eds.). (2001). *College pathways to the science education standards*. Arlington, VA: NSTA press.
- Waldvoord, M. E., Hoefnagels, M. H., Gaffin, D. D., Chumchal, M. M., & Long, D. A. (2008). An analysis of calibrated peer review (CPR) in a science lecture classroom. *Journal of College Science Teaching*, 37(4), 66-73.