# AC 2008-117: IMPLEMENTING CALIBRATED PEER REVIEW TO ENHANCE TECHNICAL CRITIQUING SKILLS IN A BIOENGINEERING LABORATORY

**Ann Saterbak, Rice University**

Ann Saterbak is Director of Laboratory Instruction and Lecturer in the Bioengineering Department at Rice University. Dr. Saterbak teaches laboratory, lecture and problem-based learning courses. She is the lead author of the textbook, Bioengineering Fundamentals, published in 2007 by Prentice Hall. She received her B.A. in Chemical Engineering and Biochemistry from Rice University in 1990 and her Ph.D. in Chemical Engineering from the University of Illinois in Urbana-Champaign in 1995.

**Tracy Volz, Rice University**

Tracy Volz is the Assistant Director and an award-winning instructor for the Cain Project in Engineering and Professional Communication at Rice University. She supports written, oral, and visual communication instruction in science and engineering courses. In addition to working with students, Dr. Volz has conducted communication seminars about oral presentations, interviewing, and technical poster design for the Texas Society of Professional Engineers and Baylor College of Medicine.

# Implementing Calibrated Peer Review™ to Enhance Technical Critiquing Skills in a Bioengineering Laboratory

## Abstract

Developed at UCLA, Calibrated Peer Review™ (CPR) is a web-based tool developed to help students improve their technical writing and critiquing skills. In 2006 and 2007 we used CPR in an upper-level tissue culture laboratory course in which students conduct viability, attachment, and proliferation assays using fibroblast cells. After completing their experiments, students use PowerPoint to construct a technical poster that illustrates their experimental methods, results, and conclusions.

For the CPR component of the assignment, students first evaluate three sample posters supplied by the instructor to calibrate their critiquing skills. After this step, students conduct a blind review of three peers' posters and then evaluate their own. During the calibration, peer critiquing and self evaluation stages, students respond to 15 statements about the quality of the posters. Eleven statements cover technical content, including succinct summary of objectives, clear experimental methods, quality of graphs, and key results interpreted in words. Three statements probe the poster's visual appeal, including appropriate size and style of font. One final statement requires a holistic evaluation of the poster. Following CPR, students turn in a revised copy of their technical poster.

In 2006, students had difficulty during the calibration phase. Following a major revision of the calibration phase in 2007, 79% of students passed all three calibration posters. Instructor, peer, and self evaluations were compared. There was a strong linear correlation between instructor evaluation and peer evaluation ($r = 0.60$, regression model ANOVA $P<0.0002$). In contrast, there were poor linear correlations between instructor and self evaluations and between peer and self evaluations ($r < 0.25$, regression model ANOVA $P>0.2$). These results suggest that students may be better able to technically evaluate others' work, rather than their own. Students perceived the peer evaluation process as generally helpful, although they noted that their peers' comments were less specific and occasionally inconsistent with their instructor's feedback. Students reported on surveys that peer evaluation was effective in helping them to recognize many facets of technical poster design, such as errors and omissions, data presentation, and technical argument. 97% of the students claimed their technical critiquing skills improved as a result of this experience. We feel that using CPR to facilitate the peer evaluation process is an effective way to enhance undergraduate engineering students' technical critiquing skills.

## Introduction to Calibrated Peer Review™ (CPR)

Developed at UCLA in 1995, CPR promotes active learning through writing and models the peer review process in science and engineering disciplines (http://cpr.molsci.ucla.edu). The National Science Foundation and the Howard Hughes Medical Institute provided initial funding for CPR, and it has been used at over 500 academic institutions.[1] According to Orville Chapman, CPR's creator, the tool enables students to "develop key skills such as abstracting, persuading

(proposals), developing arguments, describing, assessing, criticizing, analyzing, and reviewing."[2] Furthermore, the tool facilitates students' acquisition of these skills without overwhelming their instructor with additional grading.

To use CPR, students work through a five-stage process that includes Text Entry, Calibration, Review, Self-Evaluation, and Results. For the Text Entry stage, students first review instructional source material designed to help them complete a particular assignment. They then produce an original text in response to the assignment and submit it electronically using CPR's web-based interface. After they have uploaded their assignment, students begin the Calibration stage in which they respond to a set of Evaluation Statements generated by their instructor to evaluate the content and style of three sample texts. The sample texts, which are supplied and evaluated by the instructor and range in quality from poor to excellent, are specifically designed to calibrate or align students' feedback with their instructor's. In addition to receiving a grade, which is based on how closely their responses match the instructor's, students can access more detailed explanations of the instructor's response to each statement. Once students have successfully calibrated their critiquing skills, they enter the Review stage in which they use the same set of Evaluation Statements to conduct an anonymous review of three peers' texts. Finally, they assess their own text in the Self-Evaluation stage and view the results from the peer evaluation.

In addition to the student interface, CPR provides an instructor interface that is used to create and disseminate CPR assignments, source material, evaluation statements, and calibration texts. The tool also manages the anonymous peer review process, captures peer review data, assigns students grades at various stages of the CPR process, and generates reports on the results for instructors and students. For a detailed discussion of CPR's grading and reporting features, see the paper by Carlson, P.A. and F.C. Berry.[3]

**Description of the Tissue Culture Lab**

The Tissue Culture Lab (BIOE 342) is a junior-level course that is required for bioengineering majors. The course enrolls 35-50 students each spring semester. Students in the lab learn sterile technique by conducting viability, attachment and proliferation assays using fibroblast cells. The students' coursework culminates in a technical poster assignment in which they display their results from three or four of their completed assays.

The BIOE 342 course instructor developed the technical poster assignment in partnership with an instructor from the Cain Project in Engineering and Professional Communication at Rice University. The Cain Project was established in 1998 to integrate written, oral and visual communication in existing science and engineering courses. Cain Project instructors assist with course planning, communication instruction, instructional materials, and the evaluation of student work.

The technical poster assignment was first implemented in 2001. At that time students produced one poster that was graded by the course instructor. In 2002 we began refining the assignment and required students to submit a draft and a final poster, which were graded by the course instructor and the Cain Project instructor. Over the years we observed that many students were

not able to pull out the most important intellectual and methodological implications of their work. They had recurring problems constructing logical technical arguments, integrating visuals such as graphs and tables, drawing appropriate conclusions from their quantitative data, and synthesizing the results from different experiments. Therefore, we wanted to give students an opportunity to improve their skills in these areas.

**Implementation of CPR to BIOE 342**

In 2006 and 2007, we added the CPR component to the poster assignment. We introduced CPR as a vehicle by which to provide students with formative feedback on their poster drafts from their peers. Students continued to develop a poster that summarized the results of their viability, attachment and proliferation assays. They submitted a draft poster (10% of course grade), completed the CPR component (10% of course grade), and submitted a final revised poster (20% of course grade). Each poster consisted of 10-12 PowerPoint slides. While many instructors use CPR's capabilities to substitute for grading students' work themselves, we did not do so. We continued to require students to produce drafts and final posters that were graded by the course instructor and the Cain Project instructor.

The technical review is structured around 15 CPR Evaluation Statements listed in Table 1. Of those, eleven cover technical content and argument, including succinct summary of objectives, clear experimental methods, quality of graphs, and key results interpreted in words. Three statements probe the poster's visual appeal, including appropriate size and style of font. CPR forces all responses to instructor-entered evaluation statements to be on either a two-point scale (yes or no) or a three-point scale (A/B/C). The final statement requires a holistic rating between 1 and 10. We considered statements 1-14 to rate technical content and poster design and statement 15 for an overall evaluation (see Results). Seven of the statements prompt students for feedback, a written justification of the assigned score.

**Table 1.** CPR Evaluation Statements for BIOE 342 Poster Module. Statements 1-14 probe particular aspects of technical content and poster design. Statement 15 is the overall evaluation.

| | Evaluation Statements | Rating Scale | Feedback |
|---|---|---|---|
| 1 | The Objective panel adequately summarizes the scope of the work reported. | A/B/C | yes |
| 2 | The Methods panels clearly state the experimental variables. | yes/no | |
| 3 | The Methods panels explain how the key measurements were made. | yes/no | |
| 4 | The graphs and/or tables clearly present the key results. | A/B/C | yes |
| 5 | The graphs have the appropriate scale, labels and units, and "chart junk" has been eliminated. | A/B/C | yes |
| 6 | The numerical values are reported to the correct number of significant figures. | yes/no | |
| 7 | The statistics (including error bars, t-test, etc.) describe and clarify results. | A/B/C | |
| 8 | Key results are interpreted in words. | A/B/C | yes |

| 9 | The results and conclusions are supported by the data presented. | A/B/C | yes |
|---|---|---|---|
| 10 | The results from similar experiments are compared and synthesized. | yes/no | yes |
| 11 | The conclusions align with the objectives. | A/B/C | |
| 12 | The titles are persuasive and content-specific. | A/B/C | |
| 13 | The panels are concise. | A/B/C | |
| 14 | The panels are consistently formatted with the appropriate font size, style and quality. | yes/no | |
| 15 | How would you rate this text/poster? | 1-10 | yes |

Statements 1, 4, 5, 7, 8, 9, 11, 12, 13
Rating scale:　A is high or strongly agree
　　　　　　　 B is moderate or neutral
　　　　　　　 C is low or strongly disagree

Statements 2, 3, 6, 10, 14
Rating scale:　yes - agree with the statement
　　　　　　　 no - disagree with the statement

Statement 15
Rating scale:　1-10 where 10 is high and 1 is low.

When *Feedback* is required, you must enter a statement to support your rating.

Explicit instruction in poster design and CPR is given. The Cain Project instructor gives a PowerPoint presentation on "Preparing a Poster for BIOE 342." The lecture covers the fundamentals of constructing a technical argument, graphing, and includes annotated example slides illustrating common problems that have been culled from previous years' posters and revised. Following the lecture, students work in small groups and apply the lecture principles to revise 2-3 poor quality poster slides. They then use the 15 CPR evaluation statements to discuss 3-4 excellent posters designed by students in a different lab course. During this exercise, students are encouraged to comment on specific features of the posters and to ask the instructors questions (e.g. Is it appropriate to . . . ?").

In addition to preparing students for the task of constructing posters, we also train them on how to use CPR. We created handouts, including a CPR quick reference guide that explains the general purpose of each CPR stage and what action is to be taken by the student, as well as a CPR user's manual. The manual provides detailed instructions and screen shots for logging in, text entry, calibrations, reviews, self assessment, and viewing the results. In addition, we developed a CPR timeline handout to help students manage dates/times when tasks need to be completed. This is important because the dates/times are set in the software by the instructor and do not permit much flexibility. We also distribute the evaluation rubric used by the course instructor to assign grades. The evaluation rubric contains the 15 CPR evaluation statements as well as some additional items that focus on design and mechanics. In addition to these materials,

we schedule a real-time CPR demo in class and provide support for individual students who have difficulty using the interface.

CPR does not accept PowerPoint files to be uploaded as texts in the Text Entry stage, so the instructional designer developed a special procedure to circumvent this problem. Students first email their poster PowerPoint files to the instructional designer. She removes their names from the posters to protect students' anonymity during the Review stage, uploads each poster file to the Internet, and assigns a distinct url to each poster. She then sends each student an email message that contains some text and the url link to his or her poster. Once students receive this email message, they log into CPR and proceed to "Continue to Text Entry Stage." To complete the Text Entry stage, students copy and paste the contents of the entire email message into the text entry box and click "Submit Text" to complete the process.

Calibration Phase in 2006
Despite investing significant time and effort in the development of instructional materials and in training students how to use CPR, we met with a disappointing outcome in 2006. 100% of the students that semester (n=16) failed at least one calibration poster, meaning their assessment differed substantially from the instructors' evaluation. We continued with the peer and self evaluation segments; however, none of the data from 2006 are shown. We attribute this failure to our choice of calibration posters and to our CPR grading scale.

For the calibration texts, we used posters on the biomechanical properties of chicken tissues. Students in our department developed these posters, but the topic was outside the scope of the tissue culture lab. We thought the students in the course would be able to handle the complexity of the material, but they were not familiar with the terminology, experimental methods, or analysis, so they failed to distinguish between the high, medium and low quality calibration posters. We chose these posters because we were concerned that if we used calibration posters that presented the tissue culture experiments, the students would merely copy the features of the high quality poster when they revised their own posters. Normally this is not an issue in CPR because students do not see the calibration texts until after they have submitted their own texts, but we have students revise their posters for a grade after calibration, so this complicates our approach to selecting calibration texts.

We also made a mistake with regard to setting up the CPR grading template. We set the allowable deviation too narrowly between the students' responses and the instructor's responses in the Calibration stage. Specifically, students' responses to the CPR statement that requires a numeric rating could deviate by at most two points from the instructor's or the student's response was counted wrong. If students got more than 4 out of the 15 statements wrong, they failed that calibration and had to retake it.

Calibration Phase in 2007
In 2007, we generated and implemented solutions to these problems by replacing the calibration posters and adjusting our grading template. That year, 77% of the students passed all three calibration posters and 23% failed one poster and passed the other two.

To choose more appropriate posters for Calibration stage, the course instructor identified a journal article[4] that used some of the techniques the students were learning in tissue culture lab and designed calibration posters based on that article. We also piloted the calibration posters with students who took the lab in 2006. In addition, the course instructor tried to help students internalize an appropriate scale for assessment by walking the students through an entire poster and justifying her responses to the 15 CPR Evaluation Statements. However, the content of the demonstration poster was dissimilar to the course content.

We also significantly relaxed our CPR grading scale. We set the rating deviation to three points as compared to the instructor's.  Second, we allowed students to miss as many as 8 out of 15 statements on a calibration and still pass.  Finally, we encouraged students to record their calibration responses, since students are allowed two chances to pass each calibration poster.  One of the most frustrating aspects of the CPR interface from the students' perspective is that when they fail a calibration and have to retake it, CPR does not show them how they responded to the statement the first time, so many students end up repeating their errors.  Once students complete the Calibration stage, they can view and learn from the instructor's ratings of the three calibration posters and the detailed explanations that justify the instructor's response.

**Results from 2007 Implementation of CPR in BIOE 342**

Peer Evaluation
Students anonymously evaluated posters of three peers by responding to Evaluation Statements 1-14, which probed particular aspects of technical content and poster design (Table 1).   Figure 1 compares instructor evaluation to peer evaluation.  In 48% of the responses, the instructor and the peer gave the same score to a particular statement.  In 40% of the responses, the instructor rated the statement with a less positive (lower) response.  In only 12% of the cases did the peer judge a statement less favorably than the instructor (i.e., instructor rated with a more positive (higher) response than peer).  Overall, this suggests that approximately half of the peer and instructor evaluations are consistent.  When there is not agreement, the peer evaluation is less critical than the instructor evaluation in 81% of the cases.

As shown in Figure 2, there is a strong linear correlation ($r = 0.60$) between overall instructor grade (1-10 scale, with 10 high) and overall peer evaluation (Statement 15, Table 1).   Using a regression analysis, this correlation is statistically significant (ANOVA, $P<0.0002$). Despite discrepancies between peer and instructor when evaluating particular prompt statements (Fig. 1), peers were able to evaluate the overall technical quality of the poster in a manner similar to the instructor.  In other words, despite the fact that the instructor typically evaluates more critically than do students in numerical scores, the instructor and peers identify the same posters to be of high quality and the same posters to be of low quality.  Thus, we hypothesize that the process of calibration and careful peer review does illustrate student expertise in technical poster critique.
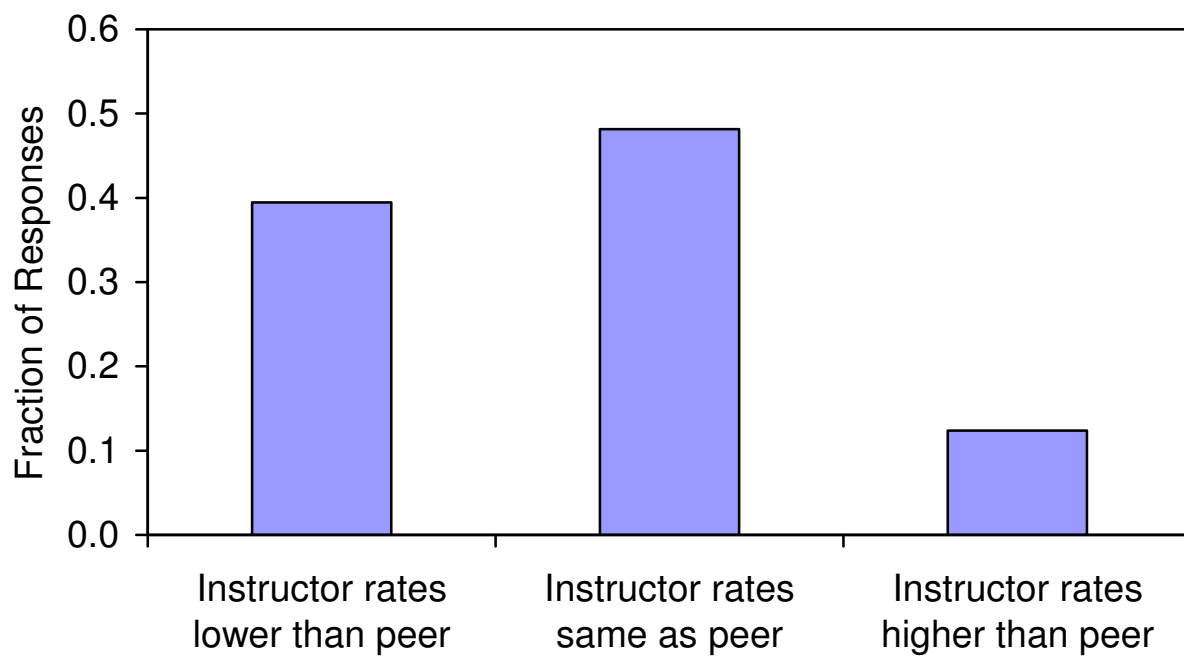
**Figure 1**. Evaluation of instructor review as compared to peer review. Statements 1-14 were considered for n = 1470 comparisons.
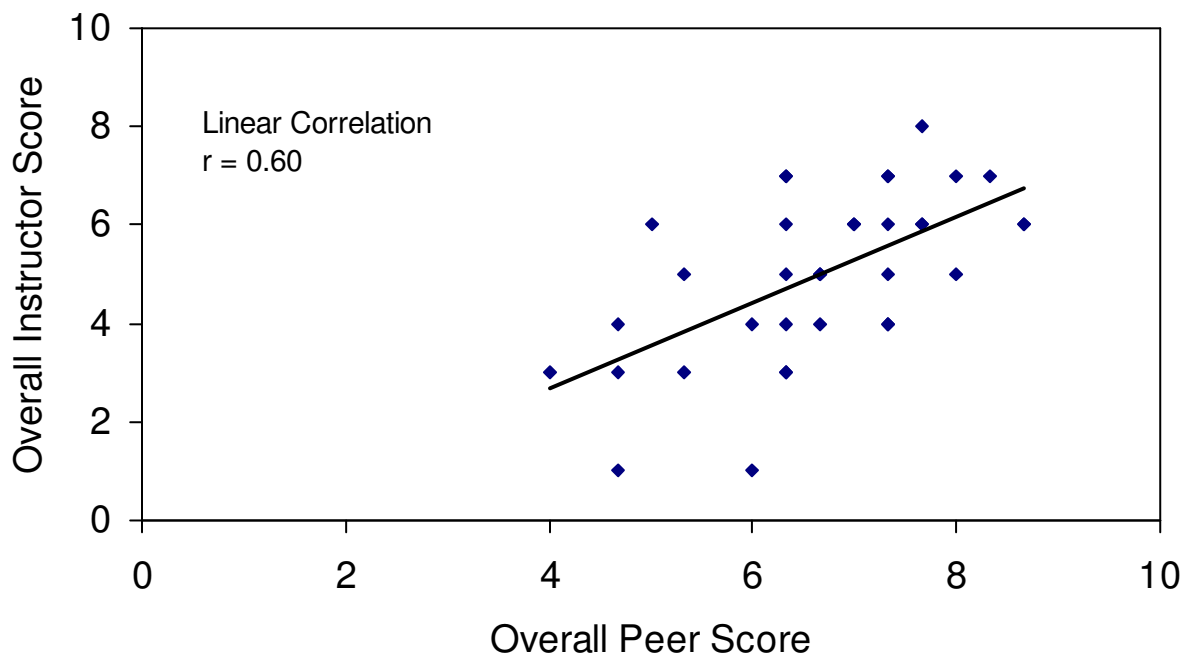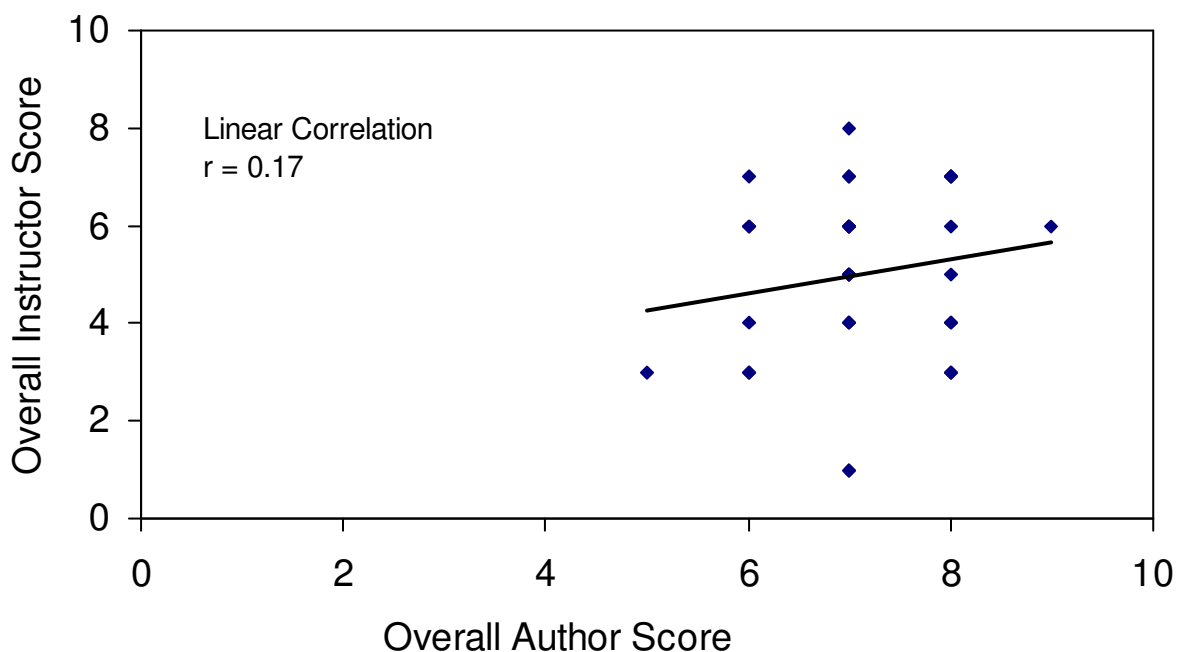


**Figure 2.** Comparison of overall instructor score to overall peer score. Peer summary score is mean of three peer evaluations for each of 35 students.

<u>Self Evaluation</u>
After completing the peer evaluation, students evaluated their own poster by responding to Evaluation Statements 1-14 (Table 1). When comparing instructor evaluation to author (or, self) evaluation, 490 comparisons were made. In 47% of the responses, the instructor and the author gave the same rating to a particular statement. In 46% of the responses, the instructor evaluated the statement with a poorer response. The author judged a statement less favorably than did the instructor in only 7% of the responses. These results are similar to those comparing instructor and peer evaluations (Fig. 1). Authors overrate themselves in 87% of the cases when there is not agreement. Thus, we found that self evaluation is less critical than peer evaluation, which is less critical than instructor evaluation.

As shown in Figure 3, there is a poor linear correlation (r = 0.17) between overall instructor grade and overall author evaluation (Statement 15, Table 1). A regression analysis shows no statistical significance (ANOVA, P>0.3). This suggests that authors were unable to evaluate the overall technical quality of their poster in a manner similar to the instructor.



**Figure 3**. Comparison of overall instructor score to overall author score.

Similar comparisons between peer evaluation and author evaluation were made with 1470 comparisons. In 56% of the responses, the peer and author gave an identical score to a particular statement. The peer rated the statement with a more negative score in 29% of the comparisons. In 15% of the cases, the peer judged a statement more favorably than did the author. The percent consistency between evaluations is the highest for this comparison (56% vs. 48% and 47%). In addition, authors overrated their work in only 67% of the cases when there was not agreement, a noticeable decrease from the other two comparisons (81% and 87%).

There is a low linear correlation (r = 0.22) between overall peer evaluation and overall author evaluation (regression ANOVA, P>0.2). Both correlations with author evaluation show an absence of a linear trend, suggesting that students do not perform well when asked to critically evaluate their own work.

Grading
The Calibrated Peer Review exercise comprised 10% of the overall BIOE 342 course grade. CPR automatically calculated a grade based on the three sections of Calibration, Review, and Self Evaluation (Table 2). All students received a perfect score on the peer review part, indicating similar peer evaluations for each particular poster. Eight or seven of the 35 students did not attain perfect scores on the calibration or self evaluation sections, respectively. The gap in calibration grades suggests additional work needs to be done by the instructor to prepare the students for calibration. The lower performance in self assessment is consistent with the data above that shows that students are more consistent when evaluating peer's work than their own.

**Table 2.** 2007 Grades

|  | Average Score | Maximum Score | # with Max. Score |
|---|---|---|---|
| Calibration | 32 | 35 | 27/35 |
| Review | 35 | 35 | 35/35 |
| Self Evaluation | 27 | 30 | 28/35 |
| Total Grade | 94 | 100 | |

Student Evaluation
Anonymous surveys were distributed to evaluate student perception of the effectiveness of CPR. Thirty-three students completed the surveys. Students were asked to evaluate whether "CPR's calibration stage fulfilled the goal of preparing me to consistently apply a set of criteria in critiquing my peers' posters" using a 4 point Likert scale. A minority of the students felt that the calibration was ineffective, with no students strongly disagreeing and 29% of the students disagreeing (n=9) with the above statement. A majority of the students felt that the calibration prepared them for peer critiquing, with 53% of the students agreeing (n=17) and 21% of the students strongly agreeing (n=7). One student who agreed noted, "Being able to see how my evaluation of the sample posters compared to Dr. Saterbak's evaluation helped me get a better sense of what she was looking for and how I should grade my peers' posters. Calibration was particularly helpful when I critiqued my own poster."

Students were asked, "Was your peers' feedback consistent with Dr. Saterbak's feedback in terms of its focus?" using a 4-point Likert scale. Three percent (n=1) felt that the feedback was inconsistent, whereas 39% (n=13) felt the feedback was somewhat consistent. 52% (n=17) deemed the feedback mostly consistent while the small fraction of 6% (n=2) felt the feedback was completely consistent. One student who responded that the feedback was somewhat consistent noted, "My peers tended to focus on stylistic issues (presentation, font, etc.) rather than the material and my logic. While presentation style does count, the subject presented should take precedence. Thus, I found myself relying much more heavily on Dr. Saterbak's comments when it came time to fix my poster." A student who responded that that the feedback was mostly consistent wrote, "Most of the key points were similar. Peers' grades tended to be a little higher

than Dr. Saterbak's scores. Peers focused less on conclusions." Overall, students noted that their peers' comments were occasionally inconsistent with and generally less specific than the instructor's feedback.

Students typically selected several responses when asked, "What insights did you gain regarding your own poster's strengths and weaknesses after you had reviewed your peers' posters? Check all that apply." Students reported that peer evaluation was effective in helping them to recognize many facets of technical poster design, such as errors and omissions (n=24), data presentation (e.g., tables and graphs) (n=22), technical argument (e.g., drawing appropriate results and conclusions from data) (n=16), precise language (n=11), organization (n=9), design (n=9), and concise language (n=9). Overall, 97% of the students claimed their technical critiquing skills improved as a result of this experience. 79% of the students felt that CPR peer review activity added to the value of BIOE 342.

**Discussion**

Instructors implement peer review with the expectation that students can be trained to provide feedback that is intellectually rigorous, thorough and consistent with faculty feedback. Falchikov and Goldfinch conducted a meta-analysis of the literature on peer review studies that included quantitative data. Their analysis of data drawn from 48 studies in various disciplines in higher education revealed that students do well when they make global judgments about their peers' work based on clear criteria.[5] The average correlation coefficient between instructor and peer scores is reported as 0.69.[5] With a correlation coefficient of 0.60 between holistic instructor and peer evaluations, our data is consistent with their work.

McCarty et al.[7] report that faculty scores on patient notes written by medical students were consistently lower than peer scores. In our work, the average instructor overall score was $5.0 \pm 1.7$, whereas overall peer and self evaluation scores were $6.6 \pm 1.2$ and $7.1 \pm 0.8$. With the instructor scores statistically significantly lower than peer evaluations (t-test, P<0.0001), it is clear that the instructor gives lower holistic scores. This data is consistent with the analysis shown in Figure 1 of Evaluation Statements 1-14 that shows that when there is not agreement between peer and instructor, the peer evaluation is less critical than the instructor evaluation in 81% of the cases. Thus, our data is consistent with that of McCarty et al.[7] She and her colleagues hypothesize that students who spend a considerable amount of time together, such as medical students in training, are more likely to review peers' work positively (even when the review is anonymous) as a result of their rapport with one another. Our discrepancy may be due to this phenomenon or may also be due to the relaxed CPR grading scale in 2007.

As stated earlier, CPR was designed so that students would move through CPR's calibration and review stages and later return to their own work at the end of the session and evaluate their texts from a more informed perspective. According to Russell,[1] "By the time [students] came to the self assessment stage [of CPR], they recognized and confronted their mistakes and accurately evaluated their own essays." Similarly, Margerum et al.'s analysis[9] suggests "students agree with their peers' review score and recognize the shortcomings of their initial work." However, this was not our experience. In the self evaluation stage of CPR, our students' consistently overrated their own posters compared to their instructor and their peers ($7.1 \pm 0.8$ compared to $5.0 \pm 1.7$

and 6.6 $\pm$ 1.2, respectively) in the holistic ratings.  This is also shown when comparing the percent of self evaluations that are in agreement and disagreement with the instructor and peers. Likewise, Prichard[6] found that "students usually rated their own essays higher than their peers' assessment of the text."  Our students appear to have been unable to achieve a critical distance on their own work. This finding contradicts our students' self-reported perception that the process of completing the calibrations and reviews enabled them to identify many types of problems, ranging from technical errors to formatting conventions, in their poster drafts.

According to survey data collected in several of studies, students report that CPR has a positive impact on their ability to write and/or to review their peers' work in a variety of contexts.[6, 7, 8] Margerum, for example, reports that more than 50% of his students agree or strongly agree that CPR improves their ability to perform technical reviews.[9]  In our case, an overwhelming 97% of students reported that their peer critiquing skills improved as a result of completing the CPR module. Our students' strong response may be the result of their having had few prior opportunities to critique peers' work, in particular technical posters. In which case, students may have concluded that any additional experience was an improvement over none.

We feel that using CPR to facilitate the peer evaluation process has improved the quality of the course. The key advantages of using CPR to conduct peer review include its ability to calibrate students' critiquing skills, preserve reviewers' anonymity, and collect student performance data. Most importantly, CPR's sequence of activities provides multiple opportunities for learning through writing. Students review six posters that they didn't author, a task they take seriously because CPR is generating a grade, and then they use that experience as well as the feedback they receive from their peers and instructors to revise their poster.

While we are generally pleased with CPR and intend to keep using it, we need to achieve greater alignment on an absolute scale between faculty and peer evaluation. To address this discrepancy, we are considering adjusting CPR's scoring template again so that students cannot pass the calibration stage so easily, and we also need to develop more instructional materials to aid students in applying the evaluation criteria in valid and reliable ways.

**Future Work**
We expect to publish a detailed analysis of specific points of agreement and disagreement on the Evaluation Statements between instructor, peer, and author in a special edition of *Across the Disciplines*.  With this analysis, we will be able to answer such questions as:
- Which particular Evaluation Statements did peers consistently overrate relative to the instructor?
- Which particular Evaluation Statements showed the most agreement between peers and instructor, or between author and instructor?
This work may reveal focus areas for further instruction in the future.

In spring 2008, the course instructor added an in-class calibration exercise.  Another calibration poster was developed based on the selected journal article.[4]  Students worked for 20 minutes in small groups to score this poster using the Evaluation Statements.  Following this exercise, the instructor gave her answers and justifications.  We hypothesize that this process will also improve success during calibration, although no data has been collected yet this semester.

Finally, we recognize that the relaxed grading scale in 2007 may have contributed to high levels of disagreement between instructor and peer and between instructor and author. In spring 2008, this grading scale was tightened to allow students to incorrectly rate only 6 out of 15 Evaluation Statements relative to the instructor on the calibration posters. Also, the allowable deviation was lowered to 2.5. We hypothesize that this change will increase the level of agreement between students and instructor.

## References

1. Russell, A. "Calibrated Peer Review ™: A Writing and Critical-Thinking Instructional Tool." (Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and mathematics (STEM) Education. Washington DC: AAAS, 2005, pp. 67-71.
2. Chapman, O. L. and M. Fiore. "The White Paper: A Description of CPR™." http://cpr.molsci.ucla.edu/cpr/resources/documents/misc/CPR_White_Paper.pdf, UCLA, 2001. Accessed 1/9/08.
3. Carlson, P.A. and F.C. Berry. "Calibrated Peer Review ™ and Assessing Learning Outcomes," *Proceedings*, *ASEE/IEEE Frontiers in Education Conference*, 2003, pp. F3E1-F3E6.
4. Gobin, A.S. and J.L. West. "Effects of Epidermal Growth Factor on Fibroblast Migration through Biomimetic Hydrogels," *Biotechnol Prog*., Vol. 19, 2003, pp. 1781-1785.
5. Falchikov, N. and J. Goldfinch. "Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks." *Review of Educational Research,* Vol. 70, 2000, pp. 287-322.
6. Prichard, J.R. "Writing to Learn: An Evaluation of the Calibrated Peer Review ™ Program in Two Neuroscience Courses" *The Journal of Undergraduate Neuroscience Education*, Vol. 4, No. 1, Fall 2005, pp. A34-A39.
7. McCarty T., M.V. Parkes, T.T. Anderson, J. Mines, B.J. Skipper, and J. Grebosky. "Improved Patient Notes from Medical Students during Web-based Teaching Using Faculty-Calibrated Peer Review and Self-Assessment," *Academic Medicine,* Vol. 80, No. 10, October 2005 Supplement, 2005, pp. S67-S70.
8. Wise, J. C. and S. Kim. "Better Understanding through Writing: Investigating Calibrated Peer Review[TM,"] *Proceedings*, *ASEE Annual Conference and Exposition*, 2004, pp. 1159-1164.
9. Margerum, L.D., M. Gulsrud, R. Manlapez, R. Rebong, and A. Love. "Application of Calibrated Peer Review (CPR) Writing Assignments to Enhance Experiments with an Environmental Chemistry Focus," *Journal of Chemical Education*, Vol. 84, No. 2, 2007, pp. 292-295.