

Students' Strengths and Weaknesses in Evaluating Technical Arguments as Revealed through Implementing Calibrated Peer Review™ in a Bioengineering Laboratory

Tracy Volz and Ann Saterbak, Rice University

Abstract: In engineering fields, students are expected to construct technical arguments that demonstrate a discipline's expected use of logic, evidence, and conventions. Many undergraduate bioengineering students struggle to enact the appropriate argument structures when they produce technical posters. To address this problem we implemented Calibrated Peer Review™ (CPR), a web-based tool, to help students improve their scientific reasoning and critiquing skills.

In 2007 bioengineering students in a tissue culture laboratory course constructed technical posters presenting their experimental methods, results and conclusions. The posters were uploaded into CPR, which permitted a highly structured approach to peer review. During the calibration, peer-critiquing and self-evaluation stages, students and the instructor responded to 15 statements about each poster's technical content and visual appeal. These statements ranged in cognitive complexity from knowledge to evaluation (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956).

Our analysis of the CPR data shows that trained peers' holistic ratings of posters are linearly correlated with instructor ratings ($r = 0.6$). In peer review, students also demonstrate expert skills, as compared with the instructor, on low-level cognitive tasks such as knowledge of material. However, students routinely overrated their peers' posters and their own as compared to the instructor on high-level cognitive tasks such as evaluation. Student self-evaluations also do not correlate well with instructor evaluations on a holistic scale ($r = 0.17$). CPR has therefore enabled us to identify where our students are having the greatest difficulty evaluating technical arguments.

Background

Rice University's Department of Bioengineering launched its undergraduate major in 1999. In designing its curriculum, the bioengineering faculty emphasized the importance of teaching students how to communicate in the discipline. To accomplish this curricular goal, members of the faculty collaborated with instructors from the Cain Project in Engineering and Professional Communication to plan and integrate written, oral and visual communication throughout the sequence of core courses required for the bioengineering major (Saterbak & Volz, 2002). The Cain Project communication instructors also

collaborated with bioengineering faculty in research and assessment projects similar to the one reported here.

In planning communication within the curriculum, the faculty chose genres that reinforced the key intellectual and cognitive activities of each course. Further, communication instruction was sequenced to reinforce students' experiences in earlier courses and to introduce more complex combinations of communication work in advanced courses. BIOE 342: Tissue Culture Laboratory was the appropriate course in which to introduce a technical poster assignment. For most of the students, a technical poster is a new genre, but it builds on the composing processes required for writing lab reports and developing slides for oral presentations, which had been part of earlier assignments in the bioengineering course sequence at Rice.

The bioengineering course instructor, Ann Saterbak, and Cain Project instructor, Tracy Volz, implemented the poster assignment in 2001, the first year that BIOE 342 was taught. Over several years of teaching this course, we recognized that BIOE 342 students were not able to identify the most important intellectual and methodological implications of their laboratory work, a failure shared by many college students. As Simpson, Layne, Godkin & Froyd (2007) note, "Students' efforts to articulate their disciplinary thinking confronts both students and faculty members with tangible evidence of confusion, misconceptions, lack of clarity" (p. 120). Our students struggled to produce the appropriate disciplinary argument structures and to support claims with quantitative data in the form of measurements, calculations and statistical analysis. Their misconceptions primarily focused on data presentation, organization, cause/effect relationships, and abstraction, more specifically, the ability to draw reasonable conclusions based on reported results. Each year we refined our instructional materials in an attempt to address these problems with visual and verbal argumentation, but the outcomes were not as good as we had expected.

In 2006 and 2007 we decided to implement Calibrated Peer Review™ (CPR) in BIOE 342. CPR is a free, web-based tool designed to support writing instruction and to facilitate peer review. We chose to implement CPR for two reasons: 1) to give bioengineering students the reinforcing experience of articulating in writing their critiques of technical posters, their peers' and their own, and 2) to improve our understanding of the problems they have in producing and evaluating technical posters.

Through a careful analysis of the data, we have realized where our students are successful and where they fall short of professional bioengineers' performances in evaluating technical arguments. Our bioengineering students' holistic ratings of their peers' posters were highly correlated to the instructor's evaluations. However, students' ability to match the instructor's evaluation on the level of individual criteria varied depending on the type of cognitive task required. These findings revealed opportunities for future study and improved instructional practice. In addition, we discovered several strengths and limitations of the CPR technology when implemented with technical posters.

This paper presents background on CPR, explains how CPR's pedagogical design serves our theoretical approach, describes the BIOE 342 poster project and CPR implementation, discusses our findings, and suggests new paths for investigation.

Calibrated Peer Review

Developed by the University of California at Los Angeles with an NSF grant, CPR is a tool that supports writing instruction and "virtual peer review" (Breuch, 2004, p. 2). CPR was originally intended to aid science educators who wanted to incorporate writing into large lecture courses without burdening the faculty with time-intensive grading. Today, CPR's use extends well beyond science educators. Over 500 institutions, ranging from K-12 programs to graduate programs, have adopted CPR. As a result, empirical evidence linking CPR to learning gains in a variety of contexts is now appearing in the literature (See, for example, Gerdeman, Russell, & Worden, 2007; Pelaez, 2002; Russell, 2005).

In brief, CPR involves four stages: composition, calibration, peer review and self-evaluation. Students first produce written material for an assignment which is uploaded into CPR. Next, students review three sample texts provided by and previously evaluated by the instructor. CPR's calibration stage trains students to apply the instructor's evaluation criteria to the samples. After they successfully complete the calibration stage, students then anonymously critique three peers' texts. Finally, students critique their own submissions, using the same criteria they applied to the calibration samples and their peers' texts.

CPR's innovative design reflects theories of active learning, Writing-across-the Curriculum (WAC) and Writing-in-the-Disciplines. The creators of CPR were guided by the assumption that students learn more when they engage actively in course concepts (Bean, 1996; Chapman & Fiore, 2001). More specifically, writing about course concepts is considered one of the most effective ways to promote learning and critical thinking (Bransford, Brown, & Cocking, 1999; Emig, 1977). CPR's pedagogical framework also represents a direct extension of WAC theory: "expository writing promotes understanding, clear writing demonstrates clear thinking, and evaluation requires higher-order, critical-thinking skills" (Russell, 2005, p. 68). Employing writing as a mode of learning should help students identify relationships, use verbal constructs to represent and integrate knowledge, and reveal gaps in their own understanding.

Of course, the students themselves, especially those in science and engineering courses, do not necessarily understand or appreciate how writing contributes to their learning. Several faculty have reported student resistance as a barrier to integrating and sustaining CPR in their courses (Furman & Robinson, 2003; Wise & Kim, 2004). Keeney-Kennicut, Gunersel, & Simpson (2008) provide a thoughtful analysis of students' positive and negative perceptions of CPR. They also describe how changes in the instructional approach, such as introducing CPR as an extension of the instructor's teaching philosophy, creating CPR-help documentation, and responding proactively to complaints about peer grades, overcame sources of resistance.

While some students may initially perceive CPR as busywork, enthusiastic proponents of CPR have argued persuasively that its highly structured and sequenced approach to peer review, which includes writing, calibrating, reviewing and reflecting, does succeed in moving students from low-level cognitive tasks such as the simple recall of facts to high-level cognitive tasks that involve analysis, synthesis and evaluation (Prichard, 2005; Russell, 2005). Carlson and Berry (2003), building on the work of learning theorists such as Emig, Vygotsky, Bloom and Perry, offer a compelling theoretical basis for CPR's ability to enhance students' critical thinking skills. In their view "the multi-staged writing workspaces in a typical CPR session encourage students to develop higher-order reasoning processes such as discerning patterns of meaning, practicing processes of inquiry, and drawing inferences from observation." They add that "practitioners who have pursued writing as a heuristic for cognition report their students are more actively engaged in learning and also find improvements in critical meta-cognitive abilities" (Carlson & Berry, 2003, p. F3E-2).

We specifically chose to implement CPR in place of traditional peer review because, as Breuch (2004) argues persuasively in *Virtual Peer Review*, electronic peer review tools such as CPR are "text-based" (p. 40). They give students crucial practice in writing about writing as compared to face-to-face, oral peer review conferences. In the process of completing a CPR module students write reviews of six texts and then reflect on the texts they have critiqued before completing a self-evaluation of their own text. CPR essentially stages learning as a recursive process of repetition and evaluation, one that is predicated upon analogical reasoning. Over time we expect our BIOE students who engage in this learning process to achieve mastery, and by mastery we mean the ability to evaluate technical arguments, their own as well as those of others, with acuity and consistency.

One of the ways this "norming" is accomplished is through CPR's calibration feature, which trains students to internalize the instructor's evaluation criteria and to apply them reliably and accurately prior to students reviewing their peers' work. In the past when we conducted traditional face-to-face peer review in class, we observed that students spent relatively little time on task, and their peer reviews of the posters focused on

low-level features (e.g., capitalization consistency), even though they were given a comprehensive poster checklist to guide their evaluations. Our experience is not uncommon. Artemeva and Logie's (2003) first attempt to implement peer review in an engineering course also resulted in "generic and shallow" feedback, and their students failed to justify their evaluations of peers' work (p. 66).

CPR's framework allows us to stretch students' critiquing skills from lower-order tasks to higher-order tasks. Furthermore, CPR assigns a grade to each student based on his or her ability to calibrate, critique, and self-evaluate, which provides a powerful incentive for many. Students have to demonstrate some degree of competency (as defined by an instructor-set grading template) in evaluating the calibration posters in order to advance to the peer review stage. If there is too much disagreement between a student's and the instructor's evaluations, the student must complete the calibration a second time. While the calibration exercise does not succeed in turning every novice reviewer into an expert purveyor of feedback, it does attempt to address the problem of "the-blind-leading-the-blind" as Carlson and Berry (2003) aptly note (p. F3E-1).

In addition to grades, CPR generates "results" for students which show them how their performance as reviewers compares to that of their peers'. We contend that one of CPR's most valuable attributes is its ability to capture and archive all of the students' responses at each stage of the process. Breuch's (2004) characterization of virtual peer review systems explains how these "fixed" comments are more "durable" because they can be referred to later by students, perhaps long after a particular assignment has been completed (p. 40).

CPR's unique data collection features captured all of the peer and self-evaluation responses associated with the BIOE 342 poster project, generating data that enabled us to probe a number of important questions. Specifically, we expected our analysis of the data to provide some insight in answering the following questions:

- Does written critique carried out through CPR lead students to evaluate others' posters holistically in the same way that the instructor rates them?
- When comparing student and instructor written critiques in CPR, do certain types of cognitive tasks routinely show agreement? Routinely show disagreement?
- How well do students self-evaluate the quality of their own posters relative to the evaluations provided by their peers and instructor?

Motivation for Technical Posters & Peer Review

Technical posters, an important genre for bioengineers to master, represent one of the primary means through which research results are disseminated to the community. Thus, it is not surprising that posters have become an increasingly popular genre taught in undergraduate science and engineering education (Kryder, 1999; Mulnix & Penhale, 1997; Vollaro, 2005; Westman, 2002). Poster assignments are a valuable addition to an engineering course because they require students to think critically about their experiments, synthesize results, distill key points, display data, and write succinctly. While many engineering genres require this set of rhetorical actions, including the integration of verbal and visual elements, we would argue that technical posters rely more extensively on visual representations that "facilitate quick visual comprehension" than some of the other genres commonly included in an undergraduate engineering curriculum (Matthews, 1990, p. 225). Furthermore, posters challenge students, more so than many other genres, to recognize and produce hierarchical, mutually reinforcing verbal/visual relationships.

Engineering students need to learn how to produce posters and to critique their technical arguments because the process of professional peer review determines whether the outcomes of research become

recognized as legitimate contributions to the community's shared knowledge. Often the presentation of results in a poster format is the "first look" that a community gets at new data. As such, meaningful, critical and balanced feedback is essential at this stage in the development of experimental work.

We therefore work to provide BIOE 342 students with insights into the processes and practices professional bioengineers use to construct and evaluate such new knowledge. Our pedagogical approach grows out of Lave and Wenger's (1991) work on situated cognition. They argue that learning is a function of the activity, context, and culture in which it occurs. Newcomers to a discourse community participate in a cognitive apprenticeship. The purpose of the cognitive apprenticeship is to introduce newcomers to the community's discursive practices and decision-making practices through authentic tasks that occur in authentic contexts through social interactions with other members of the community. Over time this socialization process contributes to the newcomers' constructions of professional identities.

In our case, the students are newcomers to the discipline of bioengineering. They are given an authentic task, to produce a poster that presents a technical argument based on primary data derived in the tissue culture lab and to review the work of their peers'. To help them understand and accomplish these tasks, we explain the purpose of peer review, introduce the poster genre, display sample posters, make explicit the tacit criteria experts use to evaluate posters, coach students on how to apply the criteria in their own evaluations, and encourage students to analyze their own performances. These activities are intended to foster independent, scientific reasoning that eventually will lead to full, professional participation in a discipline.

BIOE 342 Laboratory Course

The Tissue Culture Lab (BIOE 342) is a junior-level course that is required for bioengineering majors at Rice University, a small, private, highly selective research institution. Up to 15% of the students in BIOE 342 may be bioscience majors. The course meets three afternoons a week for four weeks. Because of its short length, students receive one credit hour for this course. The course enrolls 35-50 students each spring semester.

In BIOE 342, students learn mammalian cell culture and sterile technique. In the first week, they learn how to use light microscopes to visualize cells, correctly use micropipettes to transfer liquids, and maintain cells in culture by feeding and passaging them. During remaining weeks, students conduct six experiments using human dermal fibroblast cells. They conduct two viability assays, which enable the students to determine which cells are alive and which are dead. They conduct two cell attachment assays and explore the impact of surface treatments on cell attachment. Finally, they conduct two proliferation assays and calculate the rate of cell growth under different conditions.

During the course, students turn in carbon copies of their lab notebook pages. In addition, for each assay, they answer quantitative questions that require mathematical and/or statistical analysis of their results. Often, tasks such as graphing the results, comparing the results to a different assay, and summarizing the conclusions are required. No lab reports are required. Rather, the students' coursework culminates in a technical poster assignment in which they display their results from three or four of their completed assays (See Appendix for poster assignment).

CPR Implementation

In spring 2007, we implemented CPR as part of the technical poster assignment.^[1] There were 35 students in the class, 11 females and 24 males. We obtained IRB approval to conduct this investigation and had students sign informed consent documents.

While many instructors use CPR's capabilities to substitute for grading students' work themselves, we do not do so. Instead, we primarily use CPR to provide formative feedback. Simultaneous to the peer review of

the posters, the course instructor grades each student's draft poster (10% of course grade). The instructor also grades the final poster, which has been revised based on peer and instructor feedback (20% of course grade). In addition, students receive a grade for their CPR participation (10% of course grade). Thus, the development, critique and revision of the poster accounts for 40% of the grade in BIOE 342.

For the CPR poster project in 2007, students attended a workshop on technical poster design and a CPR tutorial. The poster design workshop built on students' prior experience of using PowerPoint and Excel to design slides for at least one presentation in a previous BIOE course taught by Saterbak. In all likelihood, many of our students also had designed posters in the past for science fairs, and some had produced posters as part of the Organic Chemistry lab at Rice, but the Organic Chemistry posters are informative rather than argumentative (They present material derived from secondary sources rather than quantitative, experimental data.). We expected the poster design workshop, as well as CPR's calibration stage, to control for some of the variation that may have been introduced by students' previous experiences, which may have involved different expectations and evaluation criteria.

From the date of the last experiment in BIOE 342, each student had one week to prepare a poster draft. The posters consisted of 10-12 PowerPoint (PPT) slides instead of one large poster landscape. We adopted this format to make it easier for peers to review individual PPT slides within the CPR interface because they would not have to resize the poster's dimensions to view specific areas more closely, as would be the case if they uploaded a large poster (e.g., 36" x 48"). In theory, this foregrounding of the constituent parts of an argument ought to make the parts and the relationships between the parts more accessible and comprehensible to peer reviewers. It should have a discriminating effect in terms of cueing a reviewer's identification of the expected categories of objectives, methods, results, etc., making it easier to evaluate.

After completing their poster drafts, students uploaded them into CPR for the calibration, peer review, and self-evaluation stages of the sequence, which occurred over a period of seven days. To complete the calibrations, students read "Effects of Epidermal Growth Factor on Fibroblast Migration through Biomimetic Hydrogels" (Gobin & West, 2003), an article the course instructor selected because it discusses several of the techniques the BIOE 342 students had learned in lab. Students then used 15 Evaluation Statements (see Table 1) to evaluate three sample posters that a Cain Project instructor had developed based on Gobin and West's paper. To pass the calibration stage, students had to match their instructor's responses to 50% of the Evaluation Statements. After completing the calibrations, students conducted blind reviews of three peers' posters and evaluated their own poster. During this time, the bioengineering instructor also graded the posters using the same Evaluation Statements. Once they finished the CPR sequence, students were encouraged to use the feedback they had received from their peers and their instructor to revise their posters. Final posters were due five days after peer and instructor feedback was provided.

Evaluation Statements Designed to Spur Scientific Reasoning

Critiquing a technical poster requires a range of cognitive tasks. The 15 Evaluation Statements (Table 1) used for the BIOE 342 poster module intentionally draw on both lower-order and higher-order skills in the cognitive domain, based on Bloom's Taxonomy (Bloom et al., 1956). Bloom's Taxonomy has been used successfully by others to design assignments in engineering courses (Irish, 1999). Bloom's model of cognition is hierarchically arranged such that higher-order tasks such as evaluation, synthesis and analysis involve, by definition, lower-order tasks such as knowledge and comprehension. In other words, one cannot evaluate an argument without knowing facts, comprehending claims, interpreting evidence, and measuring it against standards of proof.

Table 1. CPR Evaluation Statements for BIOE 342 Poster Module

When evaluating a poster, each reviewer responds to each of the 15 Evaluation Statements. The instructor created these 15 Statements, which are used by students in the role of Peer to review a classmate's poster, by

students in the role of Author to self-review, and by the Instructor to grade student work. Three different rating scales are used to respond to the Evaluation Statements:

- A (high or strongly agree), B (moderate or neutral), and C (low or strongly disagree)
- Yes (agree with the statement) or No (disagree with the statement)
- 1-10 (10 is high)

When *Feedback* is required, the reviewer must write a comment to justify his or her response; otherwise, no feedback is necessary. The final column indicates how we classified the cognitive objective of each Statement using Bloom's Taxonomy.

	Evaluation Statements	Rating Scale	Feedback	Bloom's Level
1.	The Objective panel adequately summarizes the scope of the work reported.	A/B/C	Required	Analysis
2.	The Methods panels clearly state the experimental variables.	Yes/No	--	Knowledge
3.	The Methods panels explain how the key measurements were made.	Yes/No	--	Comprehension
4.	The graphs and/or tables clearly present the key results.	A/B/C	Required	Evaluation
5.	The graphs have the appropriate scale, labels and units, and "chart junk" has been eliminated.	A/B/C	Required	Comprehension
6.	The numerical values are reported to the correct number of significant figures.	Yes/No	--	Knowledge
7.	The statistics (including error bars, t-test, etc.) describe and clarify results.	A/B/C	--	Analysis
8.	Key results are interpreted in words.	A/B/C	Required	Analysis
9.	The results and conclusions are supported by the data presented.	A/B/C	Required	Evaluation
10.	The results from similar experiments are compared and synthesized.	Yes/No	Required	Evaluation
11.	The conclusions align with the objectives.	A/BC	--	Evaluation
12.	The titles are persuasive and content-specific.	A/B/C	--	Comprehension
13.	The panels are concise.	A/B/C	--	Comprehension
14.	The panels are consistently formatted with the appropriate font size, style and quality.	Yes/No	--	Knowledge
15.	How would you rate this text/poster?	1-10	Required	Evaluation

Note that in the discussion of the Evaluation Statements that follows, we define the level of cognitive task associated with critiquing a technical poster, not with creating one. The two tasks are similar, but not identical.

Of the 15 Evaluation Statements, 2, 6, and 14 are classified as being in the knowledge domain, the lowest level of cognition. In these cases, the reviewer has to recognize that a specific piece of information has been included in the poster or that the rules of scientific notation or formatting conventions have been followed. Statements 3, 5, 12, and 13 are slightly more challenging tasks requiring comprehension. In each case, the reviewer interprets familiar material, though perhaps restructured, and then confirms or rejects the Author's explanation or approach in the absence of an explicit rule. For example, in Statement 5, the reviewer not only must remember graphing conventions but also must interpret the graph in order to assess the appropriateness of its scale. The reviewer has to understand the relationship between the variables, make inferences, and consider whether alternative configurations of the data (e.g., log versus linear scale, or scatter versus bar graph) would be more illustrative. We do not classify this statement, or any of the others in this set, as application because the reviewer is not applying an abstract principle derived from a prior experience to a new situation. Recall that the reviewer has conducted the same experiments and produced his/her own poster based on similar findings.

Statements 1, 7, and 8 involve breaking down the argument and examining its parts in relation to the whole, a cognitive task Bloom categorizes as analysis. In Statement 8, for example, the reviewer has to determine whether the author has selected the most important findings out of all possible findings and translated the meaning of that quantitative data, usually presented in the form of a graph or table, into verbal constructs. Finally, Statements 4, 9, 10, 11, and 15 require evaluation, the highest level of cognition. Note that Statement 15 requires holistic evaluation. To respond to these statements, reviewers must judge the validity and logical consistency of an argument based on its internal evidence. For example, in Statement 10 the reviewer must be able to evaluate whether the relationship between two or more experiments is clearly defined. None of the Evaluation Statements requires synthesis, putting together or combining new ideas. While synthesis is required to develop a technical poster, the reviewer does not engage in this particular type of cognitive task in the process of assessing a poster.

In summary, the Evaluation Statements developed for the CPR calibrations, peer reviews, and self-assessment range in cognitive complexity from knowledge to evaluation, based on Bloom's Taxonomy. In this group of 15 Statements, seven Statements require lower-order skills, whereas eight Statements require higher-order skills. Seven of the Statements also require student feedback. We feel this balance is appropriate for a junior-level laboratory course.

Poster Evaluation Results

Here we present the results of our implementation of Calibrated Peer Review™ in BIOE 342 in spring 2007. Thirty-five students produced technical posters about their tissue culture experiments. After uploading drafts of their posters into CPR, students completed the Calibration stage and then anonymously reviewed three peers' poster drafts using the Evaluation Statements (Table 1). When a student is reviewing another student's poster in the class, this is labeled 'Peer.' Following Peer Review, the students critiqued their own technical posters. When a student reviews himself or herself, this is labeled 'Author.' The Instructor also used the same set of Evaluation Statements when grading the posters. When the instructor is scoring the poster, this is labeled 'Instructor.'

Overall Patterns of Comparison

Table 2 reports the overall levels of agreement and disagreement between Author, Peer and Instructor responses to Evaluation Statements 1-14 for all 35 posters. The Peer and Instructor, reviewing the same poster, gave an identical rating for an Evaluation Statement in 48% of their total responses. In other words,

when comparing Peer responses to the Instructor responses for each Evaluation Statement for each poster, there was agreement in 48% of the cases.

When Peer and Instructor responses disagreed, Peer evaluations were higher or more positive than the Instructor's in 40% of those cases. An example of this would be that when evaluating the same poster, a Peer would score an A (high or strongly agree) for Evaluation Statement 1 whereas the Instructor would score a B (moderate or neutral) for the same statement. When Peer and Instructor responses disagreed, the Peer underrated or rated a statement more harshly than the Instructor in only 12% of the responses. Overall, this suggests that approximately half of the Peer and Instructor evaluations are consistent. In cases of disagreement, the Instructor's evaluations are more rigorous than Peer evaluations 81% of the time. Thus, there is a tendency for Peers to overrate relative to the Instructor, despite the calibration process.

Table 2. Peer or Author Evaluation Compared to Instructor Evaluation

All Evaluation Statements 1-14 were considered in the listed number of comparisons, *n*. Shown are the percent responses of agreement or disagreement (either overrating or underrating) when comparing Peer and Instructor responses or Author and Instructor responses. Table should be read as: Peer overrates relative to Instructor in 40% of responses. (Data published in ASEE 2008 Conference Proceedings.)

	Percent Response (%)		
	Overrates Relative to Instructor	Agrees with Instructor	Underrates Relative to Instructor
Peer (<i>n</i> = 1470 comparisons)	40	48	12
Author (<i>n</i> = 490 comparisons)	46	47	7

After completing the peer evaluation stage, students evaluated their own posters. Author responses to each of the Evaluation Statements 1-14 were then compared to the Instructor's ratings of the same poster (Table 2). The Author and Instructor gave the same rating to an Evaluation Statement in 47% of the total responses. The Author gave a higher rating than the Instructor in 46% of the responses. Thus, the percent of total responses where Author overrates relative to Instructor is basically the same (46% versus 47%) as the percent of total responses where there is agreement between Author and Instructor. The Author judged a statement less favorably than the Instructor in only 7% of the responses. When disagreement exists, Authors overrate their own poster's fulfillment of a stated criterion in 87% of the cases.

These results are similar to the comparison of Peer and Instructor responses in that students are much more likely to overrate rather than underrate relative to the Instructor. Based on these findings, Instructor evaluation is more rigorous than Peer evaluation, and Peer evaluation is more rigorous than Author (or self) evaluation.

Comparisons of Individual Evaluation Statements

We conducted a more fine-grained analysis of Peer, Author, and Instructor responses to Evaluation Statements (1-14) for each poster to identify which particular Statements generated the highest and lowest levels of agreement. Table 3 presents the Evaluation Statements that had the highest percents of agreement between Peer and Instructor (Table 3A) and between Author and Instructor (Table 3B). In each case, the four Evaluation Statements are ranked in descending frequency.

A careful look at Table 3A shows that there is agreement between Peer and Instructor in 75 of 105 comparisons (yielding 71%) for Evaluation Statement 6, which reads, "The numerical values are reported to the correct number of significant figures." The Evaluation Statement with the second most agreements was Statement 14. Note that the highest percents of responses for individual Statements showing agreement are 71% and 77% (Tables 3A and B, respectively), which are much higher than the overall levels of agreement of 48% and 47%, for the Peer/Instructor and Author/Instruction comparisons (Table 2), respectively.

Table 3. Evaluation Statements that had the most frequent agreement in responses between listed pairs of Peer and Instructor (A) or Author and Instructor (B).

The column containing # indicates the number of agreements of n possible comparisons, resulting in the given percent (%) of total responses showing agreement. The final column lists the classification of the level of cognitive task based on Bloom's Taxonomy.

A. Peer and Instructor Agreement (n = 105 comparison)

	Evaluation Statements	#	%	Bloom's Level
6.	The numerical values are reported to the correct number of significant figures.	75	71	Knowledge
14.	The panels are consistently formatted with the appropriate font size, style and quality.	74	70	Knowledge
2.	The Methods panels clearly state the experimental variables.	72	68	Knowledge
3.	The Methods panels explain how the key measurements were made.	65	62	Comprehension

B. Author and Instructor Agreement (n = 35 comparisons)

	Evaluation Statements	#	%	Bloom's Level
2.	The Methods panels clearly state the experimental variables.	27	77	Knowledge
6.	The numerical values are reported to the correct number of significant figures.	26	74	Knowledge
14.	The panels are consistently formatted with the appropriate font size, style and quality.	23	66	Knowledge
3.	The Methods panels explain how the key measurements were made.	21	60	Comprehension

Furthermore, the same set of Evaluation Statements emerges as the source of greatest agreement across all comparisons, although the order of their frequency varies slightly. In comparing Peer and Instructor responses, Evaluation Statements 6, 14, 2, and 3 present the most frequent agreements (Table 3A). In comparing Author and Instructor responses, Evaluation Statements 2, 6, 14, and 3 present the most frequent agreements (Table 3B). Similarly, in comparing Author and Peer responses (data not shown), Evaluation Statements 6, 3, 2 and 14 present the most frequent agreements.

All of these Evaluation Statements (2, 3, 6, and 14) require knowledge or comprehension, low-level cognitive tasks in Bloom's Taxonomy. For example, Statements 2 and 3 require students to verify that key words, variables, and measurements have been included in the Methods panels. Statement 6 requires the application of a general rule to a particular case. Statement 14 requires that the students recognize the appropriate application of poster design conventions. These results suggest that after CPR training most students can apply criteria consistently and accurately when it involves a low-level cognitive task.

Our analysis of the Evaluation Statements associated with the most frequent disagreements in the poster ratings given by Peer, Author, and Instructor revealed another distinct group of Statements. Table 4 lists the Evaluation Statements that produced the most overrating by the Peer relative to the Instructor (Table 4A) or Author relative to the Instructor (Table 4B). The four Evaluation Statements that occurred most often are presented in descending frequency.

Considering the four most frequent overratings on a percentage basis, there are more frequent overratings by the Author relative to the Instructor (60-83%) as compared to overrating by the Peer relative to the Instructor (49-69%). This is consistent with Table 2 which shows that Authors tend to overrate more frequently their own posters relative to Instructor (46%) than Peers overrate the same poster relative to the Instructor (40%). In addition, the percents tabulated in Table 4 are higher by 3-37%, as compared to the aggregated overratings shown in Table 2.

In Table 4A, Evaluation Statements 9, 8, 4, and 1 are the most frequently overrated by a Peer relative to the Instructor. As shown in Table 4B, Evaluation Statements 9, 8, 11, and 1 are the most frequently overrated by the Author relative to the Instructor. Three of the four Evaluation Statements (1, 8, and 9) appear on both lists.

Table 4. Evaluation Statements most frequently overrated by the Peer relative to the Instructor (A) or by the Author relative to the Instructor (B)

The column containing # is the number of overratings of n possible comparisons, resulting in the given percent (%) of responses showing overratings. The final column lists the classification of the level of cognitive task based on Bloom's Taxonomy.

A. Peer Overrates Relative to Instructor (n = 105 comparisons)

	Evaluation Statements	#	%	Bloom's Level
9.	The results and conclusions are supported by the data presented.	73	69	Evaluation
8.	Key results are interpreted in words.	65	62	Analysis
4.	The graphs and/or tables clearly present the key results.	53	50	Evaluation
1.	The Objective panel adequately summarizes the scope of the work reported.	51	49	Analysis

B. Author Overrates Relative to Instructor (n = 35 comparisons)

	Evaluation Statements	#	%	Bloom's Level
9.	The results and conclusions are supported by the data presented.	29	83	Evaluation
8.	Key results are interpreted in words.	24	69	Analysis

11.	The conclusions align with the objectives.	23	66	Evaluation
1.	The Objective panel adequately summarizes the scope of the work reported.	21	60	Analysis

All of these Evaluation Statements (1, 4, 8, 9 and 11) require students to perform high-level cognitive tasks, specifically analysis and evaluation. Based on the authors' previous experience prior to implementing CPR, Evaluation Statement 9 is the most challenging because it requires the student evaluator to interpret the data, understand the implications of the results, and to be able to evaluate whether the stated results and conclusions follow from the presented data. Not surprisingly, this Evaluation Statement ranked first in Peer and Author overrating. Several Evaluation Statements (1, 8 and 11) require the reviewer to consider parts of the poster's argument in relation to the whole in order to judge whether the material presented is logically consistent and highlights the most important findings. Statement 4 also requires evaluation because the student must judge whether the most relevant results are evidenced in figures and tables. Overall, these results indicate that most BIOE 342 students struggle with scientific reasoning, which entails interpreting quantitative data, critically evaluating verbal constructs of knowledge, identifying patterns of similarity or difference, and recognizing gaps in an argument's logic or coherence. Despite the instruction and multi-staged experience provided by CPR, BIOE 342 students continue to display novice skills in these areas.

We also analyzed the Evaluation Statements where there was the most frequent underrating of the Peer relative to the Instructor and of the Author relative to the Instructor. However, the data is not shown in the manuscript because the calculated percents of underrating were low (< 22%). This is mirrored in Table 2, which shows that the percents of underrating using the responses to Evaluation Statements 1-14 are quite low at 12% for Peer relative to Instructor and 7% for Author relative to Instructor.

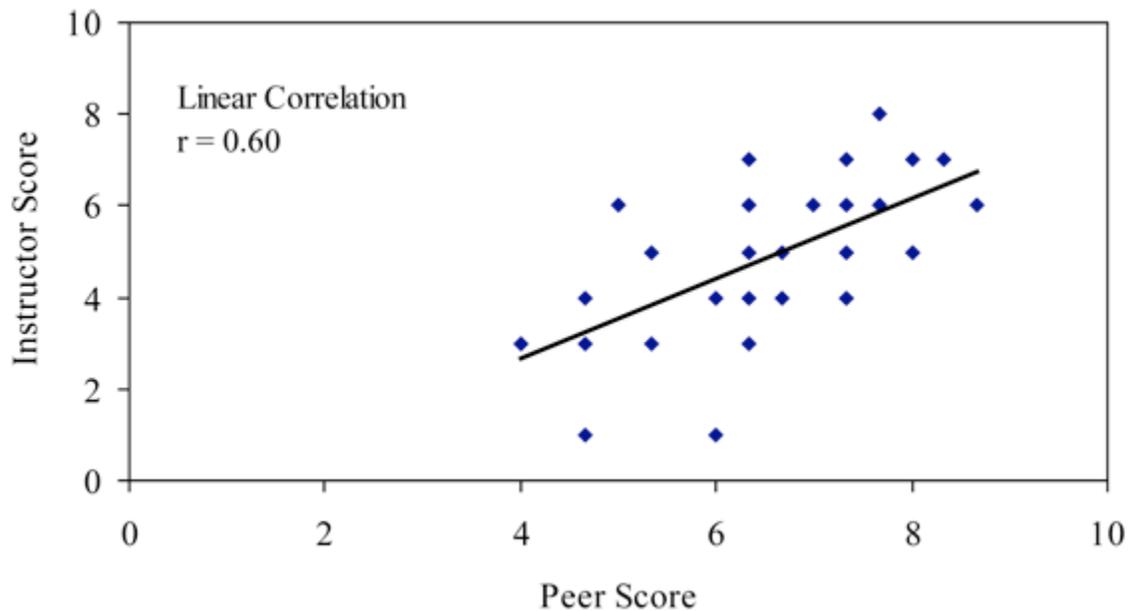
Holistic Comparisons

In contrast to Evaluation Statements 1-14, which require the reviewer to critique a particular aspect of the poster, Statement 15 requires a holistic evaluation of the poster. In addition, the rating scale of 1-10 (with 10 high) is broader. In its internal grading scheme, CPR breaks out and analyzes this Evaluation Statement separately. For these reasons, we considered Evaluation Statement 15 separately.

The Instructor's average score for Statement 15 was 5.0 ± 1.7 (mean \pm standard deviation), whereas the average Peer score was 6.6 ± 1.2 . Since the Instructor mean score is statistically significantly lower than mean Peer evaluation (t-test, $P < 0.0001$), it is clear that the Instructor gives lower holistic scores. This data is not surprising given the results in Table 2.

Because Statement 15's rating scale invites a wider range of responses, a correlation analysis could be completed. This analysis revealed a strong linear correlation ($r = 0.60$) between Instructor and Peer scores to Evaluation Statement 15 (Figure 1). Regression analysis indicates that this correlation is statistically significant (ANOVA, $P < 0.0002$). Thus, Peers were able to differentiate the overall quality of the technical posters in a manner similar to the Instructor, despite Peers' general tendency to overrate relative to the Instructor on the particular Evaluation Statements 1-14 as well as on the holistic Evaluation Statement 15. In other words, the Peers and Instructor identify the same posters to be of overall high quality and the same posters to be of overall low quality.

Figure 1. Comparison of Instructor response to Evaluation Statement 15 to Peer response to Evaluation Statement 15.



Peer score is mean of three Peer evaluations for each of 35 posters. (Data published in ASEE 2008 Conference Proceedings.)

In addition, Author and Instructor responses to Evaluation Statement 15 were analyzed and compared. In the self-evaluation stage of CPR, the Authors' average score was 7.1 ± 0.8 . Thus, Authors consistently overrated their own posters compared to the Instructor's ratings (5.0 ± 1.7) in terms of their respective holistic scores. With the Instructor scores statistically significantly lower than Author evaluations (t-test, $P < 0.0001$), it shows that the Instructor gives lower holistic scores.

A poor linear correlation ($r = 0.17$) exists between Instructor rating on Evaluation Statement 15 and Author rating in response to Evaluation Statement 15. (Data not shown; data published in ASEE 2008 Conference Proceedings.) A regression analysis indicated no statistical significance (ANOVA, $P > 0.3$). This evidence suggests that Authors failed to evaluate the overall quality of their technical poster in a manner similar to that of the Instructor. In other words, a student Author, even after Calibration and Peer Review, could not identify his/her own poster as being of high or low quality overall.

Discussion

As we stated at the outset, several years of grading bioengineering students' poster drafts convinced us that students have a difficult time producing and evaluating technical posters. We were originally motivated to study the CPR results to determine where exactly students are succeeding and failing in their ability to conduct a robust technical evaluation. Specifically, we wanted to probe the extent to which peer evaluations matched instructor evaluations and the extent to which self-evaluations matched instructor evaluations. In addition, we wanted to determine what types or levels of cognitive tasks routinely show agreement and disagreement. In the discussion, we will focus on the peer review results first, then the self evaluation results. While some similarities are noted, there are also a few distinctions.

Peer Evaluation

Our finding that Peer holistic evaluation (Evaluation Statement 15) tracks Instructor holistic evaluation is, for us, a positive result. It suggests that students were able, with some reliability, to distinguish the high and low quality posters. When considering Evaluation Statements 1-14, there is agreement between Peer and Instructor in 48% of the responses, and Peers overrate relative to Instructor in 40% of the responses. Our fine-grained analysis of the levels of agreement between Peer and Instructor on individual Evaluation Statements revealed that Peer reviewers are successful when performing lower-order cognitive tasks, such as knowledge and comprehension. In contrast, Peers overrated Statements that required higher-order cognitive tasks, such as analysis and evaluation, suggesting that students need to learn how to read and understand technical arguments in a more nuanced and critical fashion.

Because of its different rating scale, Evaluation Statement 15 was not included in the rankings that identify the statements with the most agreements (Table 3), overratings (Table 4) or underratings. However, based on mean response scores of Peers and Instructor (6.6 ± 1.2 versus 5.0 ± 1.7), it is clear that students overrated as Peers relative to the Instructor. Evaluation Statement 15 is classified as "Evaluation," since it involves making a judgment based on internal evidence and external criteria (Bloom et al., 1956, pp. 193-5). Overrating on Evaluation Statement 15 is consistent with Table 4, which shows that Peers overrate on tasks involving analysis and evaluation. Interestingly, despite the consistent overrating of Peers relative to Instructor, the bioengineering students do perform overall evaluations that highly correlate with the Instructor's. In other words, there is some evidence that students can perform critical technical reviews as Peers, a skill which is a high-level cognitive task.

Overall, these results are consistent with Falchikov and Goldfinch's (2000) impressive meta-analysis of 48 peer review studies, from which they concluded that a strong correlation exists between instructors and peers in determining global ratings, but discrepancies appear in their evaluations of specific criteria or features. Specifically, our correlation coefficient r of Peer and Instructor ratings from Evaluation Statement 15 is 0.60, which is similar to their published mean correlation of 0.69.

In our analysis of individual evaluation criteria, we found that Peers come closer to demonstrating an expert's acuity when their ratings are based on low-order cognitive tasks rather than on higher-order tasks. Peers consistently overrated aspects of the posters when they were required to analyze or evaluate specific aspects of argument. We expected the scaffolding introduced by CPR's multi-staged process to produce a better outcome, but the success of learning through analogical reasoning is contingent on students' ability to index vocabulary, conventions, concepts, patterns of organization, and question types correctly and to recall them when required. Our students are clearly quite good at indexing facts and features and applying formatting conventions. However, students continue to struggle with more sophisticated cognitive tasks. Many fail to recognize and analyze complex relationships in patterns of data. Their visual/verbal processing is not integrated. For example, students select the wrong graph or interpret it incorrectly. Some students misconstrue the definition of the relationship between variables, or they don't have a precise understanding of what the generic shape of a graph means, or they make mistakes in experimental causality.

Several possible explanations for these observations offer avenues for future research. One possible explanation for our students' inability to see links between different aspects of an argument, such as the relationship between the results and conclusions, is that posters have limited text and visuals. This lack of context ought to have been mitigated by the fact that all of the students had completed the same laboratory experiments and produced posters based on similar findings. Unfortunately, students' shared experiences may not have been sufficient to overcome the challenge of reading and assessing a technical argument in which each aspect of the argument was not only low-context but viewed in isolation. In retrospect we have speculated that the act of flipping through each poster as a series of PowerPoint slides actually may have interfered with students' ability to grasp the macro-level structure of an argument rather than making the task more manageable. This constraint, imposed by technology, may have made it more difficult for

students to perform higher-level tasks, such as judging whether the Objectives aligned with the Conclusions, because the two slides would have appeared at opposite ends of the series.

Another possibility is that Peers may have overrated posters in an effort to mask their lack of expertise in evaluating others' technical arguments. Undergraduates in science and engineering seldom have opportunities in their courses to critique one another's work (Prichard, 2005). For example, 85% of our students who completed an end-of-course survey reported that they had never before critiqued a technical poster (28 of 33 respondents). Some of these novices may have resisted the role of evaluator because they were not convinced, and rightfully so, that they had the competency or authority to critique a peer's work even after they had been provided a list of criteria and had received training in how to apply it (Herrington & Cadman, 1991; Sluijsmans, Moerkerke, van Merriënboer, & Dochy, 2001). Thus, while the BIOE 342 students have had previous opportunities to demonstrate higher-level cognitive skills, such as in a sophomore-level design project, they may be using these skills for the first time as technical reviewers. As such, their lack of experience may have hampered their effectiveness.

Finally, a combination of social pressure and camaraderie within the community also may have affected students' evaluations of their peers' performance (McCarty et al., 2005; Sadler & Good, 2006), leading Peers to overrate in general relative to the Instructor. The cohort of bioengineering majors are required to take the same course sequence, so they spend a lot of time together in class and in the lab, and many of their assignments are collaborative. Their personal interactions and the culture of the academic environment may have led them to perceive their peers to be smart and to overrate the quality of their work.

Self-Evaluation

Analysis of our bioengineering students' CPR self-evaluation data revealed that students were even less effective when assessing the quality of their own posters. Similar to Peers, Authors overrated their work relative to the Instructor in 46% of their responses to Evaluation Statements 1-14. Authors gave their own posters the highest holistic ratings (7.1 ± 0.8) as compared to Peers and Instructor (6.6 ± 1.2 and 5.0 ± 1.7 , respectively). The low correlation coefficient of $r = 0.17$ between Author and Instructor reveals that students failed to match the Instructor's holistic evaluations. It would appear that as the overrating increases, particularly on high-level tasks, students' ability to assign accurate holistic ratings decreases.

Other studies of CPR implementations that involve undergraduate students have generated mixed results with respect to self-assessment data. Prichard (2005), for example, reported that students in upper- and lower-level neuroscience courses "usually rated their own essays higher than their peers' assessment of the text," which concurs with our findings (p. A37). Gerdeman et al.'s (2007) study of learning gains associated with CPR in an introductory biology course found that lower performing students, those who initially submitted the weakest essays, showed dramatic improvements in their ability to self-assess their own texts relative to their peers' ratings. However, higher and moderate performing students showed little improvement in their ability to self-evaluate their work (See also Margerum, Gulsrud, Manlapez, Rebong & Love, 2007; Russell, 2005).

Based on our implementation of CPR, we postulate that BIOE 342 students simply do not recognize the problems in their own poster drafts. Reading and reviewing others' posters should have, by way of comparison, exposed students to the range of expertise demonstrated by their peers, which in turn should have helped them identify technical errors and problems with argumentation in their own posters. But it appears that our students "recognize other people's problems and the solutions to them much more readily than they do their own" (Gerdemen et al., 2007 quoting Rice, p. 51). Perhaps if we provide more opportunities for our students to use CPR to critique technical arguments in the future, they will be able to transfer their peer critiquing skills to their own work.

We do not know why many of our students misjudged their posters' fulfillment of the evaluation criteria. Perhaps it is because we told them that working through the CPR module would prepare them to review technical posters more accurately and consistently. In the end-of-course survey, the majority of students claimed that the peer review activity improved their ability to evaluate and critique a technical poster (Saterbak & Volz, 2008). And while the additional time on task, as well as the types and sequencing of cognitive tasks required in a CPR session, theoretically improves students' preparedness to review posters and self-evaluate, it does not succeed entirely in turning novices into experts. Yet students may have mistakenly assumed that it did.

Poor self-evaluation also may be the product of social and psychological pressure. As we mentioned earlier, most of our students have a track record of academic achievement, as measured by common metrics of student performance such as standardized test scores and grades. As such, they are psychologically invested in perpetuating the belief that they are high performers as bioengineers and that their peers are high performers, too. This perception may have affected their ratings of one another's work as well as of their own.

It is possible that students inflated their self-evaluations in an attempt to influence the instructor's grade, but students were told that their CPR grade would be based on matching their peers' reviews. Therefore, it was in their best interest to be as objective as possible in applying the criteria, and not to try to "spin" their ratings. If students did try to exploit the system in this manner, it was to no avail, because the instructor did not read the self-evaluations prior to assigning grades.

Clearly we need to investigate why our students struggle with self-assessment because it has serious implications. According to Carlson and Berry (2003), "Self-review prods meta-cognition, a capacity necessary for performance at the top levels of either the Bloom or the Perry model of student intellectual maturation" (p. F3E-3). The consequences of students not being able to self-assess effectively extend beyond a particular course or degree program. As Tan (2007) argues, the ability to self-evaluate will affect students' success as lifelong learners. They will not be able to recognize their strengths or their needs as learners much less address deficiencies and monitor their progress over time. Instead, they will make decisions about their needs that are inadequate or just plain wrong. To avoid this outcome, it is important that we not only teach students the requisite standards of performance within a disciplinary community, but also teach students how to appraise the quality of their own work and others' work accurately.

Limitations

The results of this small, classroom-based project cannot be generalized, nor can they be used to predict the outcomes of other types of engineering assignments implemented in CPR. In 2007, only the course instructor provided an expert evaluation of the students' posters, so these findings are subject to the criticism of an idiosyncratic approach to evaluation. To overcome this limitation, we need to capture and analyze poster evaluation data provided by additional bioengineering instructors and students in the future.

While our overall experience with CPR has been positive, the main barrier to using CPR is its rating scales. Students must use a narrow set of acceptable responses (e.g., many/some/none, yes/no, or A/B/C) to review texts. Only one question allows for the use of a 10-point rating scale. We have tried to compensate for this limited range of responses by requiring students to provide feedback in a textbox to justify or elaborate on many of their evaluations. Nonetheless, the narrow range of responses allowed by CPR skews the level of agreement between Author/Peer/Instructor. Specifically, the A/B/C scale forces the Evaluation Statement responses to cluster in a way they might not otherwise, if it were possible to use a 10-point scale for each question.

The persistent overrating by students relative to the Instructor may be a consequence of the CPR grading template we set for the calibration stage. The grading template allowed for a student to answer up to 50%

of the Evaluation Statements incorrectly and still pass the calibration stage. We chose to tolerate a wide margin of error due to our uncertainty in using newly developed calibration posters, which were based on a research paper. Thus, in future years, we intend to raise the calibration performance threshold with the expectation that it will result in closer alignment of student and instructor evaluations. While increasing the rigor of the calibrations ought to close the gap between expert and novice evaluations, we do not expect it to eliminate the overall pattern for the reasons discussed above.

Conclusions

Calibrated Peer Review was implemented in a junior-level Bioengineering course at Rice University. The use of CPR technology permitted a highly structured and sequenced approach to peer review, which includes writing, calibrating, reviewing and reflecting. In peer review, students show expert skills, as compared with those of the instructor, in the areas of holistic evaluation and on low-level cognitive tasks such as knowledge and comprehension of material. Students routinely overrated their peers as compared to the instructor on high-level cognitive tasks that involve analysis, synthesis and evaluation. Student self-evaluations do not correlate well with instructor evaluations. These trends in student performance have already motivated changes in our instructional approach. As a tool, CPR has been an asset to our course -- both in that it has provided a novel peer review experience for the students and in that it has offered us a unique opportunity to understand more precisely where our students are struggling in their ability to enact disciplinary arguments.

Future Paths of Investigation

CPR allowed us to identify areas of cognitive development that should be studied in the future. We need a theory of performance that reflects an understanding of the relationship between visual/verbal knowledge. Otherwise, we will fall short in changing students' behavior, and we will fail to attain our educational objectives.

Based on the CPR findings, we thought our students would benefit immediately from more preliminary work with the instructor in applying the evaluation criteria to a sample poster in class. Therefore, in the 2008 poster design/CPR workshop, we incorporated a discussion in which students and the instructor compared and justified their ratings of sample poster to help students improve their understanding of the criteria and to adjust their numeric ratings to more closely reflect the instructor's scale before students completed CPR's calibration stage. We are now analyzing the 2008 CPR data to determine whether this activity is associated with a change in student performance.

In 2008, we also added a question to our end-of-course survey to determine the extent to which viewing a poster as a series of slides within the CPR interface impedes students' ability to analyze the evidence they need to conduct accurate, expert reviews of technical arguments presented in this form. cursory review of the responses suggests that students perceive it does not interfere in a significant way.

In addition to involving students in a discussion of instructor-derived criteria and standards, we may ask students to suggest and defend their own poster evaluation criteria in the future. Falchikov and Goldfinch (2000) found a greater level of agreement between peer assessment and instructor assessment when students take responsibility for generating criteria and for determining what is significant.

Ideally our students would complete two or three iterations of the technical poster assignment using CPR, but that is simply not possible in a 4-week laboratory course. However, this fall we are working with two Biochemistry and Cell Biology faculty to implement CPR in BIOS 211: Introduction to Experimental Biosciences, which has historically enrolled nearly 25% of Rice undergraduates. Students in the course will complete three CPR assignments that focus on writing the results sections of their lab reports, and we expect this repetition to produce greater alignment across Instructor, Peer and Self-evaluations. Moreover, we

hope that these students will internalize the desired evaluation criteria and then apply them in subsequent lab courses such as BIOE 342.

Appendix: BIOE 342 Poster Assignment

Poster Summarizing Results from HDF Survival and Function *in vitro*

Task

Construct a poster to communicate effectively the viability, attachment, and/or proliferation of Human Dermal Fibroblasts (HDF) in culture conditions. Include results of three or four of your experiments. At least two of the experiments should be of the same type (i.e., attachment, proliferation, viability). The audience for your poster is your technical peers. You may include your partner's results for the cell attachment assay. You also may do so if your cells died.

Poster Requirements

The poster should contain 10-12 8.5"x11" panels (i.e. sheets of paper). Do not mount the panels on an easel or poster board.

1. Title and author (1 panel)
2. Statement of purpose/objectives (1 panel)
3. Experimental methods (1-2 panels or can be spread out over several panels)
4. Results (4-8 panels), which may include:
 - Graphs/figures (examples listed)
 - Picture of attachment patterns
 - Graph of MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] vs. cell number
 - Graph of cell # vs. time from Proliferation Assay
 - Captions summarizing results (examples listed)
 - Summary of Fibronectin (Fn) Attachment Assay
 - Effect of Fetal Bovine Serum (FBS) on cell proliferation
5. Captions comparing and contrasting results (example listed)
 - Compare Fn Attachment and Attachment Time results
6. Summary/discussion (1 panel)

Evaluation of Poster

The instructor will evaluate your poster using the Poster Evaluation Form distributed in class. Your poster should be technically strong and visually appealing.

Grades

- Draft Poster: A draft of the poster must be uploaded to CPR (Calibrated Peer Review) by the deadline. In addition, you must turn in a paper copy of your draft poster to the instructor. The draft poster will count 10% of your final course grade.
- Final Poster: Provide a paper copy of the final poster to the instructor. The final poster will count 20% of your final course grade.

References

- Artemeva, Natasha & Logie, Susan. (2003). Introducing engineering students to intellectual teamwork: The teaching and practice of peer feedback in a professional communication classroom. *Language and Learning across the Disciplines*, 6(1), 62-85.
- Bean, John C. (1996). *Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom*. San Francisco: Jossey-Bass.
- Bloom, Benjamin S., Englehart, Max D., Furst, Edward J., Hill, Walker H. & Krathwohl, David R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Co.
- Bransford, John D., Brown, Ann L. & Cocking, Rodney R. (1999). *How people learn: Brain, mind, experience and school*. Washington DC: National Academy Press.
- Bruch, Lee-Ann Kastman. (2004). *Virtual peer review: Teaching and learning about writing in online environments*. Albany, New York: State University of New York Press.
- Carlson, Patricia A. & Berry, Frederick C. (2003). Calibrated Peer Review™ and assessing learning outcomes. *Proceedings, ASEE/IEEE Frontiers in Education Conference*, F3E1-F3E6.
- Chapman, Orville L. & Fiore, Michael. (2001). The white paper: A description of CPR™. Retrieved January 9, 2008, from http://cpr.molsci.ucla.edu/cpr/resources/documents/misc/CPR_White_Paper.pdf, UCLA, 1-4.
- Emig, Janet. (1977). Writing as a mode of learning. *College Composition and Communication*, 28, 122-128.
- Falchikov, Nancy & Goldfinch, Judy. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287-322.
- Furman, Burford & Robinson, William. (2003). Improving engineering report writing with Calibrated Peer Review™. *Proceedings, ASEE/IEEE Frontiers in Education Conference*, F3E-14-15.
- Gerdeman, R. Dean, Russell, Arlene A. & Worden, Kelly J. (2007). Web-based student writing and reviewing in a large biology lecture course. *Journal of College Science Teaching*, 36(7), 46-52.
- Gobin, Andre S. & West, Jennifer L. (2003). Effects of epidermal growth factor on fibroblast migration through biomimetic hydrogels. *Biotechnology Progress*, 19, 1781-1785.
- Herrington Anne J. & Cadman, Deborah. (1991). Peer review and revising in an anthropology course: Lessons for learning. *College Composition and Communication*, 42, 184-199.
- Irish, Robert. (1999). Engineering thinking: Using Benjamin Bloom and William Perry to design assignments. *Language and Learning across the Disciplines*, 3(2), 83-102.
- Keeney-Kennicut, Wendy, Gunersel, Adalet B. & Simpson, Nancy. (2008). Overcoming resistance to a teaching innovation. *International Journal for the Scholarship of Teaching and Learning*, 2(1), 1-26.
- Kryder, LeeAnne. (1999). Mentors, models, and clients: Using the professional engineering community to identify and teach engineering genres. *IEEE Transactions on Professional Communication*, 42, 3-11.
- Lave, Jean & Wenger, Etienne. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Matthews, Diane. (1990). The scientific poster: Guidelines for effective visual communication. *Technical Communication*, 37, 225-232.
- Margerum, Lawrence D., Gulrud, Maren, Manlapez, Ronald, Rebong, Rachele & Love, Austin. (2007). Application of Calibrated Peer Review (CPR) writing assignments to enhance experiments with an environmental chemistry focus. *Journal of Chemical Education*, 84(2), 292-295.
- McCarty, Teresita, Parkes, Marie V., Anderson, Teresa T., Mines, Jan, Skipper, Betty J. & Grebosky, James. (2005). Improved patient notes from medical students during web-based teaching using faculty-Calibrated Peer Review and self-assessment. *Academic Medicine*, 80(10), October Supplement, S67-S70.
- Mulnix, Amy & Penhale, Sara. (1997). Modeling the activities of scientists. *The American Biology Teacher*, 59, 482-487.
- Pelaez, Nancy. (2002). Problem-based writing with peer review improves academic performance in physiology. *Advances in Physiology Education*, 26, 174-184.

- Prichard, J. Roxanne. (2005). Writing to learn: An evaluation of the Calibrated Peer Review™ program in two neuroscience courses. *The Journal of Undergraduate Neuroscience Education*, 4(1), A34-A39.
- Russell, Arlene. (2005). 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, 67-71.
- Sadler, Philip M. & Good, Eddie. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.
- Saterbak, Ann & Volz, Tracy. (2002). Integrating communication into the biomedical engineering curriculum. *Proceedings, International Conference of the IEEE Engineering in Medicine and Biology*, 3, 2660-2661.
- Saterbak, Ann & Volz, Tracy. (2008). Implementing Calibrated Peer Review™ to enhance technical critiquing skills in a bioengineering laboratory. *Proceedings, American Society of Engineering Education Conference*. Reference number: 2008-117.
- Simpson, Nancy, Layne, Jean, Godkin, Blake & Froyd, Jeff. (2006). Faculty development through student learning initiatives: Lessons learned. In Douglas Reimondo Robertson & Linda B. Nilson (Eds.), *To Improve the Academy*, Vol. 25. (pp. 109-122). Bolton, Massachusetts: Anker Pub Co., Inc.
- Sluijsmans, Dominique M., Moerkerke, George, van Merriënboer, Jeroen G. & Dochy, Filip J. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation*, 27, 153-173.
- Tan, Kelvin. (2007). Conceptions of self-assessment: What is needed for long-term learning? In David Boud and Nancy Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term*. (pp. 114-127). New York: Routledge.
- Vollaro, Mary B. (2005). More than science fair fun: Poster session as an experiential learning activity in the classroom. *Proceedings, American Society of Engineering Education Annual Conference and Exposition*, 10537-10545.
- Westman, Jennifer J. (2002). Introduction to scientific research: Research experiences for undergraduates. *IEEE's Proceedings of the American Control Conference*, 2, 1092-6.
- Wise, John C. & Kim, Seong. (2004). Better understanding through writing: Investigating Calibrated Peer Review™. *Proceedings, ASEE 2004 Annual Conference and Exposition*, 1159-1164.

Notes

[1] We also implemented CPR in 2006 with mixed results. (For a more complete description of our CPR implementation, including our preliminary work in 2006, see Saterbak & Volz, 2008.)

Acknowledgements

We would like to thank Liz Eich, Mary Purugganan, and Sharon Gibson-Mainka for supporting the CPR poster assignment in BIOE 342. We also wish to thank Jan Hewitt and Linda Driskill for their comments on our manuscript.

Contact Information

Tracy Volz, PhD
Brown School of Engineering MS-364
P.O. Box 1892
Houston, Texas 77251-1892
Voice: (713) 348-6243
Email: tmvolz@rice.edu

Ann Saterbak, PhD
Dept. of Bioengineering MS-142

P.O. Box 1892
Houston, Texas 77251-1892
Voice: (713) 348-6243
Email: saterbak@rice.edu

Complete APA Citation

Volz, Tracy, & Saterbak, Ann. (2009, January 19). Students' strengths and weaknesses in evaluating technical arguments as revealed through implementing Calibrated Peer Review™ in a bioengineering laboratory. [Special issue on Writing Technologies and Writing Across the Curriculum] *Across the Disciplines*, 6. Retrieved from https://wac.colostate.edu/docs/atd/technologies/volz_saterbak.pdf