

AC 2007-2484: A WEB-BASED TOOL FOR IMPLEMENTING PEER REVIEW

Patricia Carlson, Rose-Hulman Institute of Technology

Dr. Patricia A. Carlson received the BS from the College of William and Mary in 1968 and the MS and Ph.D. degrees from Duke University in 1969 and 1973 respectively. Currently Dr. Carlson is Professor of American Literature and Director of PRISM, Department of Humanities and Social Science, Rose-Hulman Institute of Technology.

Frederick Berry, Rose-Hulman Institute of Technology

Dr. Frederick C. Berry received the BS, MS, and DE degrees from Louisiana Tech University in 1981, 1983, and 1988 respectively. He taught in the Electrical Engineering Department at Louisiana Tech University from 1982-1995. Currently Dr. Berry is Professor and Head of the Electrical and Computer Engineering Department at Rose-Hulman Institute of Technology.

A Web-based Tool for Implementing Peer-Review

1.0 Introduction

Over the last several years, engineering education has been in the process of reinventing itself. This unprecedented change is but a part of reform-driven shifts in teaching goals, pedagogical methods, and course content taking place across the nation at all levels of instruction. One facet of this change in engineering education has been a renewed emphasis on student teams and on student-provided formative feedback within an assessment process anchored in learning outcomes.

The authors report on the integration of Calibrated Peer Review™ (CPR™) – a web-delivered student feedback tool – used in three courses at Rose-Hulman Institute of Technology. Since academic year 2002, the authors have developed course activities that highlight writing and peer evaluation as central components of

- RH131 (Rhetoric and Composition): An introductory composition course required of all students at this college of engineering.
- ECE 361 (Engineering Practice): A sophomore-level course covering project design specifications, team roles, effective conduct of team meetings, written and oral communication skills, ethics and professionalism, completion of team project(s).
- ECE 362 (Principles of Design): A junior-level course covering conceptual design, scheduling, project management, business plan, market survey, and budgeting that culminates in a written proposal and oral presentation requesting funds for development of a product.

We report on the results of our using this method of giving student-generated feedback, which has been successfully used by hundreds of engineering students over the course of several years at RHIT. The paper and the poster examine CPR™'s approach to implementing peer review and how these methods measure up to generalized expectations for computer-mediated collaborative assessment.

2.0 Peer Review, Outcomes Assessment, and Formative Feedback

Reform-driven engineering education incorporates various types of collaborative learning experiences. Such pedagogy yields a number of gains for modern engineering education. Peer review is an especially fruitful technique, whose instructional outcomes should:

- Enhance students' meta-cognitive abilities in a complex process by fostering, higher-order activities, such as those represented by the upper levels of Bloom's Taxonomy [1]
- Encourage students to move toward mature, professional behaviors, such as the progression outlined by the Perry Model [9]. Within this framework, the student progressively moves from depending on external, "teacher-centered" authority to a more self-assured ability to reconcile multiple perspectives, to tolerate ambiguity, and to reflect on the process itself (meta-cognition).

- Encourage students to develop the social skills needed to work with a team through the sharing of ideas, the ability to provide meaningful, constructive feedback, and the ability to accept peer critiques.

Unfortunately, integrating effective peer-review sessions into a course requires much effort on the part of the instructor. Karen Spear [13] enumerates several of the pitfalls associated with peer-review of writing, no matter what the course content:

- Confused expectations about the group's purpose and the individual's role in it;
- Inability to read group members' texts analytically;
- Misperceptions about the nature of revision and of writing as a process;
- Failure to work collaboratively with group members;
- Failure to monitor and maintain group activities (pp. 17-18)

In addition to this question of student competency, other frequently heard detractors include amount of class time consumed by the process, low student motivation translating into an inability to stay on task, and difficulties in capturing learning outcomes for a process that may be conceptualized as highly subjective and basically open-ended.

3.0 What is Calibrated Peer Review™?

Full descriptions of CPR™ appear on the system's website (<http://cpr.molsci.ucla.edu/>). Additionally, we have provided more detailed examination of CPR™ as a learning tool elsewhere [3 - 5]. Here we include only a composite of the system's features and functions in order to guide our discussion of CPR™ in relation to emerging understandings of computer-mediated peer review (CMPR).

First introduced in 1999, CPR™ is a component of a large-scale, National Science Foundation-supported project led by a team of educators at UCLA to develop a completely digitized, network-delivered Molecular Science Curriculum. The fully integrated CPR™ contains an assignment authoring tool for custom crafting of writing tasks and a library of edited assignments contributed by instructors from varied disciplines and institutions. Currently hosted at UCLA, the system draws from the model of manuscript submission and peer review in the conduct of scientific inquiry [6].

Four structured workspaces perform in tandem to create a series of activities that reflect modern pedagogical strategies for using writing in the learning process. A separate instructor interface and student interface provide customized reports on performance for individual assignments (see Figure 1). These separate student and instructor reports are data- and information- rich. In the body of this paper, we examine how these reports constitute a

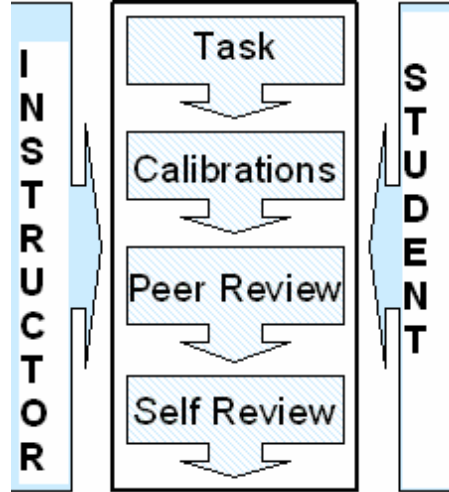


Figure 1: A Dynamic, Multi-Staged Learning Environment

- **Task:** Students are presented with a challenging writing task, with guiding questions to act as scaffolding for the demanding cognitive activities. Students compose using a word processor, but upload the finished text as an HTML file. Graphics and all tables are supported in the upload through standard HTML tags.
- **Calibration:** Students read through three “benchmark” samples and assign each a score based on a series of evaluative questions (a rubric). Students are then given a “reliability index” from 1 to 6, based on their demonstrated competency in these exercises. This segment mitigates the common objection to peer review in the undergraduate classroom: that the experience reduces itself to the-blind- leading-the-blind.
- **Peer Review:** After becoming a “trained-reader” – and being assigned a credibility weighting – students read and provide written feedback on three anonymous peer essays using the same rubric as used in the calibrations. Students also assign each essay a holistic score from 1 to 10.
- **Self-Assessment:** As a final activity, students evaluate their own essay. As with calibration and peer review, students use the same “rubric” (set of performance standards for the task). Having “trained” on benchmark samples, and then applied their expertise in evaluating peer text, students now engage in a reflective, final activity by assessing their own submission. Students are encouraged at this time to make comments to themselves (and also available to the instructor) that capture the evolving insights they have gained in the previous two segments. They are also invited to reflect on whether they have gained a deeper level of understanding for the assignment and its outcomes.

CPR captures and stores performance data for each student at crucial junctures in the peer review process. Observations from seven categories measure dimensions of the process of learning for individual students or for aggregates of students. Table A explains what these seven measurements are, how they are represented, and why they are useful as formative feedback, both for instructors and for students.

Table A: Useful Data for Analysis

Workspace	Data Measurement	Use
Text Rating	Quality of the Artifact	Expressed as a number from 1 (low) to 10 (high); this score is the average of the holistic evaluation made by each of the peer reviewers.
Calibration	% Content	The items within the rubric are divided into two broad categories, broadly characterized as “content” and as “style.” These scores are reported as percentages of correct answers to a series of questions.
	% Style	
	Average Deviation	The average difference between student holistic ratings and answer-key ratings for all three samples given in the calibration workspace.
	RCI (Reader Competency Index)	Based on overall performance in the calibration exercise, students are given a “credibility” score of from 1 (lowest) to 6 (highest).
Peer Review	Average Deviation	The average difference between student ratings of peers’ text and ratings from two other student reviewers of the same text within the same CPR session.
Self-Review	Average Deviation	The average difference between student self-rating and the average of ratings from all three peer reviewers.

In addition to empirical data, the CPR™ system also stores (and displays on request) all the peer-provided, text-based commentary for each student, from each session. Samples of these student-written critiques are given below. Based on the instructor’s preference, CPR™ results may be used either as formative or as summative. In other words, CPR lends itself to mediating revision or to providing a student-awarded final grade. The implementation is up to the instructor. Complex algorithms in the system collect observations from thirteen separate learning events within the process and tally a score based on the traditional 100 points. The professor sets the points awarded for each major event and the total may be used as a grade. Students can easily examine the data on their performance summation to determine which aspects of their participation (text, calibration, providing peer reviews, or self-reflection) need improvement.

4.0 How Does CPR Measure Up to General Expectations for CMPR?

Peer review in the classroom has become increasingly popular as an instructional device. Computer-mediated peer review (CMPR) offers advantages on several dimensions: convenience, uniformity of delivery, data collection, and student satisfaction. While emerging technologies make asynchronous peer editing more engaging and available, the literature is only now beginning to identify the core dimensions for designing or evaluating online peer review systems.

Elsewhere we have examined how CPR™’s *in-situ* data collection provides evidence for learning outcomes [5]. Here, we consider how CPR™ addresses the most common performance criteria for collaborative learning involving peer-review. Our observations are clustered around the central questions for the efficacy of peer review as a pedagogical device as suggested by Edward F. Gehringer in a paper given at the 2001 ASEE Annual Conference [8].

4.1 Assigning Collaborators in a Peer-Review Process

Having students critically analyze their classmates’ work assumes that all students have at least some base-line ability to effectively judge the assignment under consideration and to offer useful comments for improvement. Sadly, experienced instructors know that this is not the case. Students come to the socio-cognitive task of peer reviewing with widely varying skills, and both symmetrical and asymmetrical skill levels pairings can result in difficulties. In a traditional classroom, countering these pitfalls requires a significant investment of instructor preparation and use of precious class time.

4.1.1 How are peer reviewers assigned to each submission? CPR™ assigns student reviewers (three for each submission) on a random basis. While students are not paired, the process is reciprocal in that all students are writers and all students are reviewers. Authors and reviewers are anonymous to students, but in the instructor’s report, each individual is identified. By exporting the CPR data to a spreadsheet, instructors can look for patterns of performance based on individuals or aggregates, within or across assignments.

The student evaluators are not “uninformed” by the time they reach the peer-review workspace. As indicated in Figure 2, the student has gone through the calibration training session, in which she examines three “benchmark” responses to the assignment (of high, medium, and low quality) and provides two types of reader response: (1) answers a set of analytical questions assessing fine-grained aspects of each piece and (2) assigns a holistic score (1-10) for the entire submission. Students receive system-generated guidance during the calibration when their answers are too far from the norm established by the instructor.

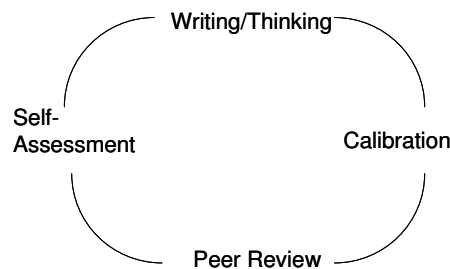


Figure 2: Sequence of Events in Calibrated Peer Review’s Four Structured Workspaces

The assumption for CPR™ is that the calibration portion of the exercise helps to smooth out the differences in the reviewers’ abilities. Also alleviating the problems of varying skills levels, each student receives a rating of her achievement during the calibration (from 1 to 6). This number is used to “weight” the impact of her scoring of classmates’ submissions during the peer review.

4.1.2 *Is there evidence indicating that “assignment” is an important variable in the process?* As the teaching profession gains experience with CMPR, nuanced questions about what really matters in the process will be more closely examined. At this point, we offer the following section to demonstrate that CPR’s data collection capacity makes the system an excellent research tool for contributing to the knowledge base on how best to design and implement CMPR.

An indication that CPR™ addresses the varying skills issue, both on an experiential level and on an empirical level, we have found that CPR™ sometimes helps to “level the playing field” for the less talented students. However, the determining factor is clearly the quality and clarity of the calibration exercise as much – if not more – than the fact that it is implemented within a computer-mediated environment.

For example, using the data collected during the six CPR sessions in ECE 362, we calculated the standard deviation mean for holistic scores given by student reviewers for each assignment. Offered in the Spring Quarter of 2004, this two section course had a total of 55 students (12 women and 43 men). Both sections had the same instructor.

To this analysis of mean standard deviation on peer reviews for all six assignments (CPR sessions), we added a measurement of how students’ initial aptitude affects performance. We did a median split on all six sessions, using the mean of holistic scores received on the text submission. In other words, we used performance on the assignment to – somewhat arbitrarily -- separate our population into high and low ability groupings. We then plotted the average standard deviation from the norm on peer-reviews performed by each group for each of the six sessions. The x-axis gives the session or assignment number, and the y-axis gives the value range for the mean standard deviation. The lower the standard deviation from the norm the better the student’s performance in the peer review component.

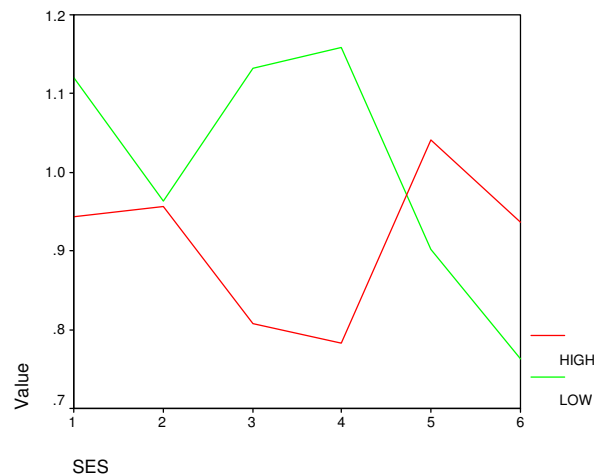


Figure 3: Average Deviation in Peer Review Plotted Across Six Sessions for -- Median Split Using Text Rate Score

As can be seen in Figure 3, these preliminary results do not indicate a clear and consistent negative slope for either ability group (as determined by the median split). Nevertheless, we found three observations of interest.

- First, of all the assignments, the writing prompt in session #2 seems to dampen whatever effect initial competency had between the two aptitude groups.
- Second, session #4 clearly treated the two aptitude groups very differently (indicating an aptitude-treatment interaction worth investigating). The variance between the high aptitude's and the lower aptitude's ability to apply what was learned in the calibration training caused us to re-examine the calibration examples used in session #4.
- Third, the training in session #6 appears to have worked well for both groups, but seems to have been especially effective with the lower aptitude group.

We examined the calibration training in each of these assignments to determine how best to revise the materials for upcoming courses. Demonstrating that an assignment was effective for all students -- and was especially effective with a low-ability group -- indicates robust instruction.

4.2 Reliability and Validity of Peer-Provided Feedback/Responses

Many things can go wrong in a situation where students are asked to judge their fellow students. While students seldom frame their concerns using the terms “reliability” and “validity,” their reluctance to accept peer critiques often centers on questions of credibility and confidence. Mistrust of classroom colleagues as a source of useful guidance results in such pitfalls as

- Discomfort or defensiveness during the peer-review process exhibited by some students, and
- Less-than-candid behaviors evolving during the collaboration process (such as mutual admiration societies)

Despite its enormous potential, peer-review suffers from the common criticism that – without labor-intensive scaffolding and time-consuming mentoring – the process frequently reduces itself to “the blind-leading-the-blind.”

Calibrated Peer Review™ has several features that enhance both reliability and validity in student peer responses. The carefully orchestrated calibration phase – as described above – improves a student's ability to recognize both rhetorical and content requirements for the specific assignment. Additionally, the calibration exercise results in a confidence rating (from 1 to 6) for each student, which is used in determining how much that student's response affects a fellow student's final score. This alleviates much of the fear of being “down graded” by a commentator who does not know good from bad. Furthermore, internal evidence for the predictive value of the confidence rating score (CRS) for the accuracy of text ratings can be calculated on a class aggregate basis by looking at the correlation between a low CRS numbers and high standard deviations in holistic text scoring.

Providing a reader confidence index is only one of the algorithms that compute an individual reviewer's accuracy of assessment during the review process. Three other methods for guidance and consolidation will be mentioned briefly.

- Scaffolding in the Form of a Richly Developed Rubric: The authoring functions of CPR™ and the visualization of a complex, multi-staged process made possible by the four structured workspaces help instructors to set learning goals and to devise scaffolding by which novice learners can accomplish those goals. Of these CPR™ scaffolding devices, the most powerful is the rubric. Instructors unpack an assignment through a cognitive task analysis and accompanying outcomes, and then ask students to rate how well their colleagues have met these individualized performance standards.

Repetition and reinforcement are central to the CPR™ learning process. Students work with this same rubric in all three evaluative phases of the session: calibration, peer review, and self-reflection. CPR™ allows for flexibility both in uploaded artifact (which could be textual or visual) and the attendant assessment items. Thus, instructors have few limitations in crafting their rubric. For example, the item being peer critiqued could be a PowerPoint® presentation or a digitized video of an oral report. The rubric could require a “yes/no” response with a written explanation or it could present a Likert scale and ask for the degree to which a sample has meet a standard, with a written explanation.

- Incentives for Accurately Applying Concepts from the Calibration Phase: The composite session report provides students with feedback on their accuracy of their evaluations, as indicated by how far they were from the numerical norm. (Instructors can set the degree of difficulty for this measurement using the low, medium, or high settings.) This provides a motivation for students who might otherwise complete the peer evaluations in a casual or random pattern.
- Richly Contextualized Modeling of Written Feedback: Learning how to give constructive commentary challenges many students new to peer reviewing. In general, their written advice may be too skimpy to be useful or they tend to make the same sketchy and generic comments over and over. CPR's fundamental strength accrues from its power to model for students good examples of mature commentary.

4.3 Student Perceptions of Peer-Review

While educational theorists and practitioners alike promote the value of having students learn from other students, more telling are studies that done on how students view the process. Simkin & Ramarapu report that students feel comfortable with peer-reviews of their written materials provided a procedure to promote fairness and rich response has been instituted [12]. Rafiq & Fullerton found that students teams who kept highly-focused diaries on members' performance were more engaged in all aspects of the course [10]. In general, students are least satisfied with peer evaluation methods that produce superficial, vague, glib, or harsh commentary. In general,

the literature reinforces the notion that if students get/give only sketchy feedback, they view the process as worthless “busy work.”

Our students usually find the first CPR™ session challenging. Seldom – especially in an engineering course – have they been held accountable for the process of writing to this degree. (Even in classes where instructors require peer critiques of documents, it is difficult either to mentor or to monitor students at this fine-grained a level.) However, our experiences show that over the ten-week quarter, students come to value the CPR™ experience.

In ECE 362, for example, each CPR™ assignment highlights a critical component of the larger, final proposal; therefore, students learn the iterative nature of composing a quality piece of writing. They also come to trust their peers’ judgment and to value the guidance they receive from fellow students. Anecdotal evidence suggests that almost all students are positive by the end of the course. Even those who struggle will admit that they better understand how to write a project proposal and that many of the fine-grained composition skills they have learned come from scrutinizing the submissions of fellow students. Also, a simple comparison between the final proposals from the most recent offering of ECE 362 (Spring Quarter, 2006) and final proposals from the same course as taught in years without CPR™ shows dramatic improvement.

4.4 Quantitative and Qualitative Feedback to Mediate Improvement

CPR provides learners with two types of detailed formative assessment: quantitative (the numerical ratings given by peers on various aspects of the assignment) and qualitative (the narrative comment used to justify particular scores). We briefly examine each.

4.4.1 Quantitative Ratings -- The complex algorithms of the CPR™ system provide each student not only with numerical evaluations of her submitted text but also an indication of how well she evaluated the work of others. As mentioned previously, the student also receives a reader confidence index score indicating how qualified she is as a reviewer, based on performance during the training (or “calibrating”) segment of the session. These are unique features of CPR™; the system is able to indicate both the reliability and the validity of individual’s peer-review contributions. Such feedback motivates students to improve their scores, and alleviates the anxiety expressed by students that “unqualified” raters are negatively affecting their grades.

Figure 4 shows how performance information is summarized and returned to students. Students see how well they performed in evaluating three of their peers (top level of the interface). They also get more detail on each of the performance criteria contained in the rubric (five for this particular assignment). The student also sees the summative, holistic score (from 1-10) provided by each of her three peer reviewers. Confidence levels for the three readers are indicated in the weights applied to each reviewer’s score. At the bottom of the results screen, students see how performance measures in each of the four workspaces add up to a final score (based on 100 points).

We have found that students respond well to this fine-grained numerical feedback. They are able to track their progress on each component in the writing/reviewing process, both for individual assignments and across multiple CPR sessions.

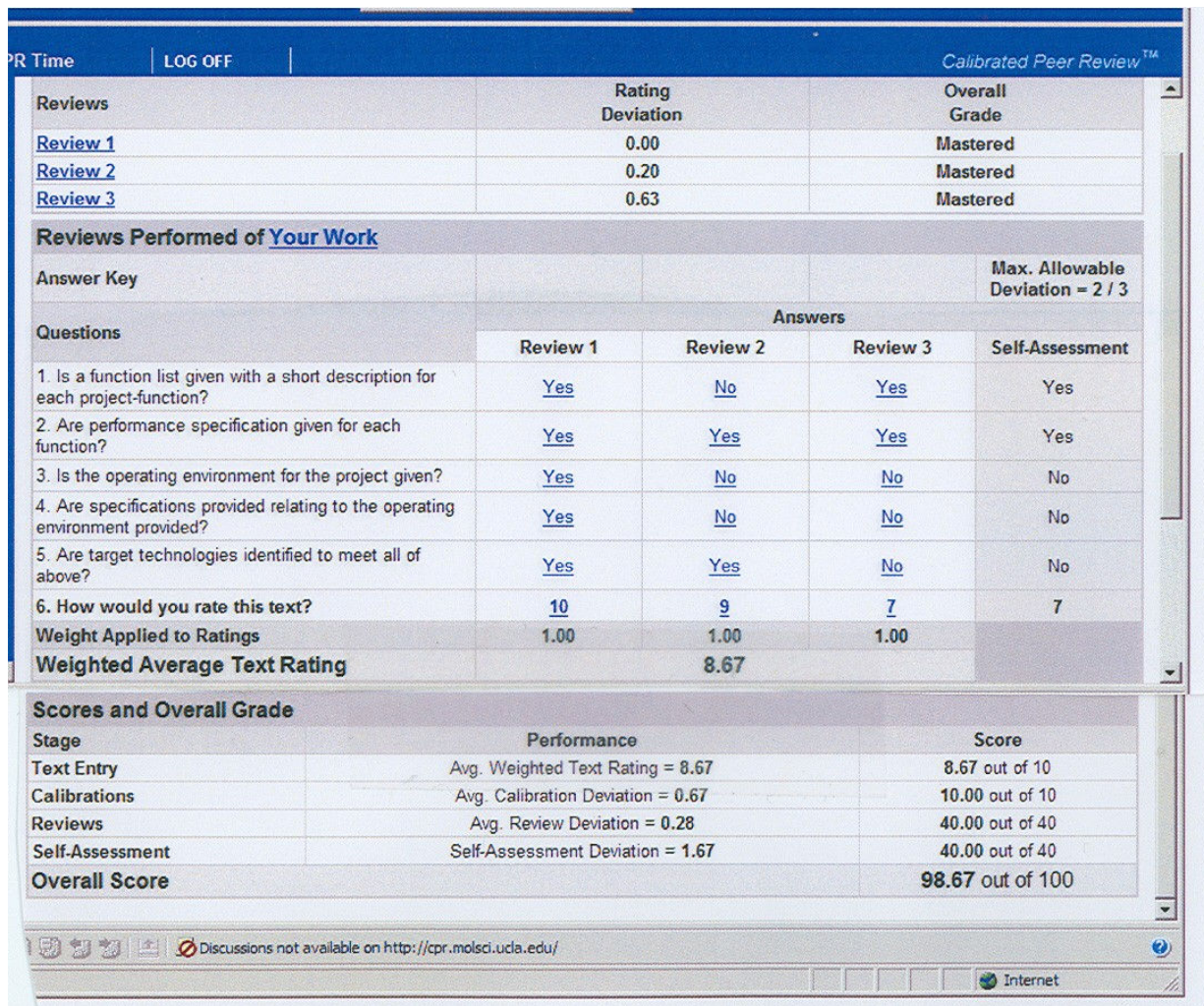


Figure 4: Sample CPR Student Session Results Screen

4.4.2 *Qualitative Comments*: As motivating as the quantitative feedback is, the advice and observations provided by peers becomes invaluable in guiding students' growth and improvement [7]. By clicking on the response to each rubric item (see Figure 4), the student author is taken to what the reviewer said to justify the numerical evaluation.

We have found that by carefully crafting the performance questions in the rubric, we are able to elicit peer commentary exhibiting characteristics of the three higher levels of Bloom's Taxonomy:

- Analysis: Breaking down objects or ideas into simpler parts and seeing how the parts relate and are organized
- Synthesis: Rearranging component ideas into a new whole

- Evaluation: Making judgments based on internal evidence and external criteria.

Table B provides some sample student commentary taken from assignment #3: Product Design Specifications, for ECE 362. They illustrate how the guided-inductive nature of the CPR™ embedded rubric elicits focused evaluation. Students move beyond description/explanation (a lower level cognitive skill) toward using discerning judgment in order to help a colleague solve a problem.

TABLE B: EXAMPLES OF QUALITATIVE FEEDBACK FOR PDS

RUBRIC ITEM	STUDENT SAMPLE RESPONSE (from the range of papers in this session)
(1) Is a function list given with a short description of each project-function?	NO -- Explanation: Components are given, but specific functions are not described. This needs to be organized into a list so it is easier to read. There is a lot of information here that is not needed in the PDS.
(2) Are performance specifications given for each function?	YES: Explanation: some of the functions don't need specs, but for example "disc title memory" - 24 titles/ 8 letters, that's a good example of performance spec for a function.
(3) Is the operating environment for the project given?	YES -- Explanation: May want to expand on existing cars, what type of temperatures will your device need to endure? Having a different type of taillight means the lights will have to endure some sort of corrosive tests.
(4) Are specifications provided relating to the operating environment?	NO -- Explanation: In a car there are government regulations even relating to a non safety-critical system such as taillights. Look at the Federal Motor Vehicle Safety Standards. The one that should apply to brake lamps is FMVSS 108: Lamps, Reflective Devices and Associated Equipment. Also, you will need to probably have some sort of failsafe scheme to assure that a failure is detected. This is not as important as your device can not by itself cause a potential hazard, but if your device says there is no problem and doesn't warn the driver because of a device failure, that seems like you are opening yourself up to problems such as a lawsuit.
(5) Are target technologies identified to meet all of the above?	NO -- Explanation: There are a number of different transducers. Is this going to be one that measures brake pedal travel, or is this going to be brake force. Also, the brake travel is not a linear function when compared to the amount of actual decel on the car. From car to car this can vary greatly. Also, there is a large discontinuity at the point of booster run out. Another point is a number of cars already have some sort of brake pedal travel sensor, especially ones with newer brake control systems. Also, what type of technology is going to be used as an "electronic tape measure"? You can use Ladar, Radar, or other systems. These can have advantages and disadvantages. Some of the laser based systems have problems in weather.
(6) How would you rate this text? (on a scale of 1 to 10)	5 -- Explanation: It is very brief and difficult to pull out information quickly. You might want to reformat it, add more detail, and discuss your environmental factors.

6.0 In Conclusion

Although the “peer review” of student writing has demonstrated impressive learning gains in the teaching of mathematics and sciences, relatively few engineering programs have adopted the approach, potentially because of the drawbacks mentioned above.

Calibrated Peer Review™ has the characteristics of what an educational technologist would call a “cognitive tool,” a representational device aiding learners to enact more powerful strategies for accomplishing problem-solving than possible without the heuristics embedded in the device. In practice, cognitive tools occur in a number of different domains. Their form spans a gamut from simple (but powerful) mnemonic devices for extending the working memory (such as a grocery list) to rich, highly complex models that foster expert behaviors by invoking appropriate strategies, conserving and allocating mental energies, and orchestrating steps in staged problem-solving techniques.

We have described some advantages of using CPR™ as a platform for integrating peer review into engineering education. However, any instructor considering CPR™ for course adoption will also want to know about such pragmatic issues as ease of use and overhead for instructors and return on class time invested.

- Overhead for Instructors: Authoring a CPR™ session is labor-intensive for the first couple of times. However, once the instructor builds up some expertise and a small library of adaptable assignments, the task becomes easier. Depending on the individual and the complexity of the assignment, a session may take four to five hours to prepare. In our experience, the return on investment comes in being able to treat written work seriously without burying oneself in stacks of grading or returning documents with copious commentary, which students may all-too easily ignore, misinterpret, or misplace.
- Investment of Course Time: Each instructor decides how much emphasis can (or should) be given to CMPR assignments in a given course. A few caveats are appropriate here. First CPR™ sessions work best if the writing assignment is relatively short and compact (say, two, three, or four paragraphs). Second, the assignment should involve problem solving, critical thinking, or concept formation. Furthermore, the objectives of the assignment should be well-formulated and clearly reflected throughout the CPR™ session. In our several years of combine experience with the system, we have found CPR™ most appropriate for *drafting* key components of longer exercises. On average, students in the pilot study worked directly with the CPR environment for no more than two or three hours per week. Given the return in learning, we believe the time is extremely productive.

We have found no system available today that duplicates the following powerful features of CPR™ as a complex, highly orchestrated cognitive tool for mediating peer review:

- The writing/review/revision process – perceived as opaque and arbitrary by many students – is reified into a set of understandable “state transitions” through the robust “thinking frame.”

- Students are trained on what to look for in evaluating the specific assignment, improving the quality of student feedback, enhancing learner confidence, and promoting deep engagement in the process.
- After evaluating peer contributions, students can consolidate and apply their gains by reflecting on their own submission (in light of insights gained from examining materials from fellow students).
- The system handles all logistical issues: materials are stored and distributed in keeping with the sequence of the “thinking frame”; student submissions are randomly distributed and reviews are anonymous; powerful, built-in algorithms measure levels of performance at key stages in the process.

Works Cited

- [1] Bloom, B., Englehart, M. Furst, E., Hill, W., & D. Krathwohl. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: Longmans-Green, 1956.
- [2] Breuch, L-A. K. (2004). *Virtual Peer Review: Teaching and Learning about Writing in Online Environments*. Albany, NY: State University of New York Press.
- [3] Carlson, P., Berry, F., & Voltmer, D. (2003). “Using Calibrated Peer Review to Mediate Writing and to Assess Instructional Outcomes,” *Proceedings*, American Society for Engineering Education Annual Conference, Session 2793.
- [4] Carlson, P. & Berry, F. (2003). Calibrated Peer Review and assessing learning outcomes, *Proceedings*, IEEE/ASEE Frontiers in Education.
- [5] Carlson, P. A. & Berry, F. C. (2005). Calibrated Peer Review: A tool for assessing the process as well as the product in learning outcomes,” *Proceedings*, ASEE Conference.
- [6] Chapman, O. L, & Fiore, M. A. (2000). Calibrated Peer Review. *Journal of Interactive Instructional Development*, 12(3), 11-15.
- [7] Cho, K. (2004). *When multi-peers give better advice than an expert: The type and impact of feedback given by students and an expert on student writing*. Unpublished doctoral dissertation, University of Pittsburgh.
- [8] Gehringer, E. F. (2001). Assignment and quality control of peer reviewers, *Proceedings*, ASEE Annual Conference, Session 3230.
- [9] Pavelich, M.J. & Moore, W.S., (1993). Measuring maturing rates of engineering students using the Perry Model,” *Proceedings*, Frontiers in Education Conference, pp. 461-466.
- [10] Rafiq, Y. & Fullerton, H. “Peer Assessment of Group Projects in Civil Engineering,” *Assessment and Evaluation in Higher Education*, Vol. 21, No. 1, pp. 69-81, 1996.
- [11] Schiver, K. (1992). Teaching writers to anticipate readers’ needs: A classroom-evaluated pedagogy. *Written Communication*, 9, 179-208.
- [12] Simkin, M.G. & Ramarpu, N.K. (1997), Student perceptions of the peer review process in student writing projects, *Journal of Technical Writing and Communication*, Vol. 27, No. 3, pp. 249-263, 1997.
- [13] Spear, K. I. (1988). *Sharing writing: Peer response groups in English classes*. Portsmouth, NH: Boynton/Cook.

Acknowledgement:

Much of the work was accomplished with the support of NSF grants #9980867 – “Using Writing as a Heuristic for Scientific Problem Solving in a Computer-mediated Environment” and #0404923 – “Writing for Learning and Assessment in Engineering Design Courses.” Any

opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.