Evaluating Calibrated Peer Review™:
Implementation at Rose-Hulman Institute of Technology

Carslon, P. A., Berry, F. C. and Voltmer, D. (2003, June). "Using Calibrated Peer Review™ to Mediate Writing and to Assess Instructional Outcomes." Proceedings of the 2003 American Society For Engineering Education Annual Conference and Exposition, session 2793. Nashville. 13 pp.

Carslon, P. A. and Berry, F. C. (2003, November). "Calibrated Peer Review™ and Assessing Learning Outcomes." 33rd ASEE/IEEE Frontiers in Education Conference, session F3E. Boulder. 6 pp.

Carslon, P. A. and Berry, F. C. (2005, March). "Calibrated Peer Review (CPR): A tool for integrating meaningful writing assignments into technical courses." Proceedings of the NCIIA 9th Annual Meeting. San Diego. 8 pp.

Carslon, P. A. and Berry, F. C. (2005, June). "Calibrated Peer Review: A Tool for Assessing Process as Well as the Product of Learning Outcomes." Proceedings of the 2005 American Society For Engineering Education Annual Conference and Exposition, session 1981. Portland. 21 pp.

Berry, F. C. and Carlson, P.A. (2007, June). "ABET Assessment Using Calibrated Peer Review." Proceedings of the 2007 American Society For Engineering Education Annual Conference and Exposition, session AC 2007-186. Honolulu. 7 pp.

Carslon, P. A. and Berry, F. C. (2007, June). "A Web-based Tool for Implementing Peer-Review." Proceedings of the 2007 American Society For Engineering Education Annual Conference and Exposition, session AC 2007-2484. Honolulu. 14 pp.

Berry, F. C., Carslon, P. A., Eccles, W., Ferguson, B., Moore, D., Radu, M., Schumacher, T., Voltmer, D., Yoder, M. and Wheeler, E. (2007, June). "An Undergraduate, Entrepreneurial Design Sequence: A Decade Of Development And Success." Proceedings of the 2007 American Society For Engineering Education Annual Conference and Exposition, session AC 2007-1753. Honolulu. 15 pp.

Calibrated Peer Review™ (CPR): A Brief History

Calibrated Peer Review™ (CPR) is an asynchronous, interactive web-based program designed to engage students in writing activities that promote learning. The software was originally developed as part of the Molecular Science Project in the late 1990s by Orville Chapman at the University of California, Los Angeles. His goal was to create a discipline-independent, computer-based learning tool that would facilitate writing-across-the-curriculum

pedagogies by virtue of alleviating the burden of correcting and grading written assignments (Chapman and Fiore, 2001; Russell, 2004; UCLA and USC, 2005; UCLA, 2001).

CPR has grown substantially in recognition and popularity since its public launch in 2001 (Russell, 2004, June 11). At that time, it was implemented among 101 partner institutions and 250 courses. The assignment library housed 175 classroom-ready modules and was accessed by 16,000 student accounts (Russell, 2005, p. 69). By 2004, the user-base had grown to 500 institutions, 1,900 courses and 72,000 student accounts, and the assignment library contained 1,275 CPR modules (p. 69). A recent count in 2007 indicates an exponential surge in growth to roughly 900 institutions and roughly 135,000 students (Russell and Fiore, 2007).

CPR has even generated "spin-offs." At the American Society for Engineering Education Annual Conference in 2005 and again in 2007, two separate groups presented curricular strategies that claimed to be based on CPR methodology. In 2005, Ohland, et al., reported addressing the problem of subjectivity in peer evaluations by choosing "to train students how to use the instrument, thereby calibrating their responses in much the same way as is used in Calibrated Peer Review" (p. 3). In 2007, McStravik and O'Malley reported on their capstone curriculum design, which "implemented a new peer feedback activity for presentations and the initial design plan written report. These activities are based on the Calibrated Peer Review method but do not use the online software" (p. 5)

A Question of Efficacy

Chapman and Fiore and Russell, CPR's respective developers and principal investigator, claim it improves critical thinking and knowledge retention across all disciplines, regardless of how learning outcomes are assessed. They also assert that this learning outcome can be achieved without imposing onto faculty large burdens of additional time or money. (Chapman and Fiore,

2001; Russell, June 2004; UCLA, 2001; Gerdeman, Russell and Worden, 2007). CPR is a grant-funded resource and free for both instructors and students. This enables faculty who could not otherwise write their own programs to benefit from computer-assisted learning tools. Faculty can use any of the assignments in the shared library or author their own. Authoring an assignment from scratch requires an investment of several hours; however, assignment monitoring and grading are accomplished via the software (UCLA and USC, 2005; Russell, March/April 2004). Authoring a lesson takes considerably less time than writing a program from scratch, even for accomplished programmers (Russell, March/April 2004).

Regardless of its popularity, the question arises—does CPR really work? This is more difficult to determine, in large part because it's so new. Even early adopters are still in the early stages of figuring out how to implement it effectively and integrate it into their teaching strategies. Thus far, the two dozen-or-so published findings report positive learning outcomes. However, it should be noted that many of the published results are based on qualitative findings, such as student surveys of satisfaction or preference. (Likkel, 2007; Margerum, et al., 2007; Prichard, 2005). Other studies compare before and after written work that is not blinded, i.e., the evaluator is the instructor of the course and therefore aware of which work was completed after exposure to CPR (Pelaez, 2002; Plutsky and Wilson, 2004; McCarty, et al., 2005). Most of the analyses consist of simple calculations, such as percent change in performance or proportion of overall results.

Although still scarce, rigorous analyses are beginning to show up in conferences and publications. In a recent publication, Gerdeman, Russell and Worden conducted statistical analyses on the effect of CPR on freshman/sophomore biology students over three different

quarter terms [n=1,330] (2007, p. 48). They reported that CPR did improve student essays, with

the greatest improvements coming from students who had the lowest initial skill levels (p.51).

Also lacking are details of the pedagogical context in which CPR is being implemented.

In other words, most of the results reflect studies of CPR *in vitro*, like a stem cell in a petri dish,

separate from an overall curricular strategy. This kind of isolated analysis is fundamental to

scientific research and critical for studying the effects of an individual mechanism or

functionality; the isolation reduces the number of environmental or subjective variables that may

influence the final results. Then again, since CPR is not supposed to be used as a stand-alone

substitute for teacher-student interactions, the results of an isolated study might not reflect its

true value as a component of an integrated learning environment.

Evaluating CPR at Rose-Hulman Institute of Technology

With these beginning efforts in mind, the seven essays reviewed here yield valuable

information about how CPR operates within a fully integrated pedagogical strategy. Collectively,

the papers provide a comprehensive view of one particular CPR implementation, from start to

current status. The frequency of publication—seven publications over four years—provides fine-

grained progress reports of the foundational theory and initial implementation, as well as

ongoing refinements.

The primary authors are Frederick C. Berry, professor of engineering and Patricia A.

Carlson, professor of history and writing, from the Rose-Hulman Institute of Technology

(RHIT). This ABET-accredited[1] undergraduate university focuses on engineering, science and

mathematics. Their educational philosophy is that "writing is important to engineering—as it is

to any knowledge worker—because the act of placing ideas into language mediates higher-order

---

[1] Accreditation Board for Engineering and Technology (ABET) is an accreditation organization for post-secondary
education programs in science, computing, engineering, and technology.

intellectual activities that are foundational to critical thinking and multi-staged problem solving"

(Carlson and Berry, June 2003).

Using Calibrated Peer Review™ to Mediate Writing and to Assess Instructional Outcomes (June 2003)

Calibrated Peer Review™ and Assessing Learning Outcomes (November 2003)

The first pair of papers was published in 2003, the same year in which CPR was

implemented as a component of an end-to-end, four-course engineering specialty program called

Design Sequence (Carlson and Berry, March 2005; Berry, et al., 2007). Both papers describe

how CPR operates to create an interactive learning environment and how it mediates the peer

review process. The papers also describe learning goals for each of the fours stages, or

workspaces—task, calibration, peer-review, self-assessment—that comprise a complete CPR

module. These descriptions are essentially the same as those presented by Chapman, Fiore and

Russell (2001, 2004, 2007), so there is little reason to expand on them here.[2]

Carlson and Berry propose a two-part thesis against which CPR will be evaluated at

RHIT. The first part is that CPR engages students in a multi-staged writing activity that develops

higher-order reasoning processes—such as discerning patterns of meaning, practicing processes

of inquiry and drawing inferences from observations. The educational theory postulated for how

CPR produces these learning outcomes is based on Bloom's Taxonomy of Cognitive Objectives

and Perry's Model of Affective Growth. Bloom's cognitive taxonomy describes the learning

process as beginning with passive acceptance of knowledge and shifting to active inquiry and

meta-analysis. Perry's affective model focuses on decision-making; starting with decisions

influenced by absolute standards and outside authority and shifting to decisions that

accommodate ambiguity and rely on internal authority.

---

[2] One point of note, they diagram the software engine with descriptions of the software processes running in each stage, a geeky extra that is omitted in subsequent papers and presentations.

This is a departure from Chapman and Russell who define the program as an emulation of the scientific-research process without referencing a psychological model. Graham writes, "Peer review has a prominent role in science. Anonymous peer review is thus the model on which we built Calibrated Peer Review™" (Chapman and Fiore, 2001). Russell notes "Scientists and engineers do research and write and peer-review research manuscripts.…writing with anonymous peer review is thus the model for CPR" (Russell, April 2004).

The second part of Carlson and Berry's thesis makes the novel premise that CPR also works as an *in situ* assessment tool. In this context, they will study CPR to determine if it provides quantifiable measurements of ABET Engineering Criteria 2000 of academic competencies. Measuring competency for the ambiguous section 3, item g., or EC3(g), "ability to communicate effectively" is particularly challenging. In this regard, they believe that CPR has the potential to refine and improve the definition of EC3(g) "beyond the level of a cliché" (p.2). Furthermore, they suggest that analyzing collected data using standard and specialized reduction protocols should enable assignment authors to "deconstruct" what is happening during the interactions between student, assignment and program, and furthermore, enable them to make corrective adjustments to improve the learning gains.

"Using Calibrated Peer Review™ to Mediate Writing and to Assess Instructional Outcomes" takes their theory about learning outcomes a step further by proposing that CPR "encourages students to develop strategies that will endure even in the absence of the computer-mediated learning environment" (pp. 6-7). This paper suggests the insertion of cognitive "markers" when authoring assignments to track whether a module is producing the desired learning outcome (pp. 9-10). It should be noted that these markers are not mentioned in later papers, so perhaps they were determined not to be useful in actual practice.

"Calibrated Peer Review™ and Assessing Learning Outcomes" suggests that CPR data collection might be used to create a portfolio of student work over an entire academic career. This "electronic portfolio" could provide both a snapshot for a particular point in time, as well as cumulative view of progress over time. The information could be used as either an assessment or teaching tool, depending on how it is applied (pp. 5-6).

Calibrated Peer Review (CPR): A tool for integrating meaningful writing assignments into technical courses (March 2005)

Two years later, Carlson and Berry provide a more specific look at how CPR is being implemented in a technical-writing course called Engineering Practice. This paper is primarily an expository "how-to" for faculty who might be considering implementing CPR into their curricular strategy. The paper provides a description of course requirements and detailed breakdown of eight CPR modules, out of a total of 12, assigned for the course.

The chronology of assignments shows how the modules were designed to work incrementally and cumulatively. Assignments progress from theory to application, as well as from simple to complex writing. Two sets of modules on report writing are repeated—modules 4 and 7 on Product Design Specifications, and 5 and 6 on Project Technical Description. This introduces students to the use of revision as a writing tool, and is enforced by increasing the complexity and difficulty of the second assignment of each set.

"Calibrated Peer Review" A Tool for Assessing the Process as Well as the Product in Learning Outcomes (June 2005)

Here we get the first glimpse of whether the promise of CPR is proving to be true at RHIT. Two years after implementation, RHIT is still firmly committed to the educational philosophy "writing as an analog for thinking in engineering design" (p.1) and the premise that

"through the vehicle of CPR, we were able to implement assignments that fully utilize the writing across curriculum (WAC) pedagogy" (p.6).

The paper includes a more refined diagram and explanation of the four workspaces of the CPR module. It offers a quick comparison to web-delivered courseware such as Angel™ or WebCT™. There is a short section with practical advice on ease of use, return on time investment and student reactions. The discussion of the statistical analyses is preceded by a description of the various data logs and reports (pp.3-8).

As indicated by the title, the focus of this paper is on the program's capabilities as an assessment tool. It is essentially a progress report on the second part of their original implementation thesis—the *in situ* data collected by CPR can be analyzed statistically to yield quantifiable results. The goal was to see if the quantitative results supported the qualitative and simpler analyses reported in other papers. In short, does CPR improve learning and more importantly, can we deconstruct what is happening?

From a sample size of 55 students under the same instructor, data was collected from six CPR sessions assigned over a 10-week academic quarter. Overall, statistical analyses of the data supported the premise that CPR is an effective teaching tool that improves learning. The descriptive and regression analyses indicated that students who learned to recognize rhetorical features during the calibration stage were able to translate that knowledge into accurate peer assessments (pp. 9-11, 13). Correlation of means and calculations of variances helped identify fine-grained relationships between calibration questions and corresponding aspects of the holistic peer review (pp.11-13). Analysis across assignments, which was not done statistically, focused on the electronic portfolio functionality of the data collected (pp. 14-15).

Carlson and Berry conclude by expanding their original two-prong thesis regarding CPR's utility into four areas: 1) a writing tool for learning complex thinking and behavior skills; 2) a computer-mediated tool that facilitates advanced socio-cognitive development; 3) a teaching tool that addresses challenging pedagogies, namely engineering design; 4) an assessment tool for ABET-style accreditation (pp. 16-17).

It is important to point out that although concept 2 fits into the theories of Bloom and Perry, it is probably the least well supported by the data presented in this paper. The results illustrate student learning outcomes entirely within the context of CPR assignments of one specific course. The claim of advanced socio-cognitive development would have to be substantiated by studying how the students behave when they are in a different course and not using CPR. Thus far, Carlson and Berry have yet to produce evidence that fully supports their assertion that CPR "internalize[s] strategies for later performance of the same or similar tasks, without the presence of the technology" (p. 2). They are certainly aware of this shortcoming, as the footnote on page 8 indicates they are conducting another study "to determine if CPR fosters cognitive development as defined by the six stages in Bloom's Taxonomy of Cognitive Objectives."

ABET Assessment Using Calibrated Peer Review (2007)

This short paper furthers the learning assessment discussion by exploring how CPR might be used to improve reporting on academic competencies, specifically ABET engineering criterion section 3, item g, or ABET g, "an ability to communicate effectively" (p.1); ABET j, "a knowledge of contemporary issues; and ABET i, "a recognition of the need for, and an ability to engage in life-long learning" (pp.5-6).

CPR assignments from 54 students in the same course were analyzed and the results showed promise as an assessment tool for ABET competencies. However, Berry and Carlson felt that the results could be improved with modifications to lesson design. The paper promised further studies in this area.

A Web-based Tool for Implementing Peer Review (2007)

With this paper, Carlson and Berry nicely bookend the assessment capabilities of CPR with an analysis of how it teaches higher-order reasoning, the first part of their original thesis. They point out the merits of its asynchronous delivery and 24/7 availability for out-of-class assignments. They also discuss how it alleviates several pitfalls associated with peer-review in general: confused expectations of review process and purpose, variability in analytical abilities of reviewers, misunderstanding of writing and revision process, failure to work collaboratively, lack of monitoring and mediation and poor returns on investment of class time (p.2).

Because the peer reviews are randomly and blindly assigned by the CPR software, student evaluators are unlikely to introduce personal bias towards well- or ill-favored classmates. The calibration workspace helps standardize competencies across the group prior to the actual peer reviews. At the end of the calibration process, each student evaluator receives a rating score that is used to weight the impact of their reporting during the peer review workspace (p.5). Quality of the peer reviews can be further monitored by examining the correlation between predictive values such as confidence rating score (CRS) and standard deviation values generated during the later workspaces (p. 7).

The quality of the assignment authoring is crucial to the success of the learning experience. A well-authored assignment provides both repetition and reinforcement of the learning objectives. With this goal in mind, the paper recommends that careful attention be paid

to the three phases: calibration, peer review and self-reflection. Analysis of the mean standard deviation is suggested for pinpointing potential areas of assignment modification within individual CPR modules (pp. 6-7).

Student evaluators are required to provide both quantitative and qualitative feedback during the peer and self reviews. In return, they receive feedback on their abilities as a student evaluator and the quality of their assignment as rated by their peers. The feedback serves to motivate students to improve their scores by developing better critical-thinking capabilities, both as evaluators and writers. The qualitative assessments, which are the written commentaries, provide further opportunities for developing higher-order analysis processes (p. 11).

In conclusion, Carlson and Berry restate their thesis that CPR is a powerful tool for teaching complex problem-solving and advanced cognitive development. They also conclude that as a computer-mediated peer-review tool, CPR surmounts the pitfalls of traditional peer-review. They do acknowledge that a significant amount of time and effort must be initially invested, namely during assignment authoring, to maximize the benefit received from a CPR implementation. They also recommend that CPR works best for short writing assignments. For longer written documents, CPR should be used to help students draft individual components of the finished piece.

An Undergraduate, Entrepreneurial Design Sequence: A Decade of Development and Success (June 2007)

This paper describes the end-to-end curricular framework of which CPR is an actively contributing component. The paper covers details of individual courses: context in the overall sequence, topics and content, learning objectives and schedules. Design is a relatively new specialty for the Engineering Department at RHIT and the paper includes discussions of what worked, what didn't, student evaluations and faculty response. It highlights the careful planning

that went into how and where CPR was incorporated. Four years from the initial implementation, this paper provides an excellent time-stamped and contextual reference for the CPR project thus far. This program overview should be particularly useful for faculty considering how to include CPR in their curricular strategies and pedagogical goals.

What Can We Learn From RHIT?

To begin with, it would be a stretch to claim that RHIT research has provided enough evidence of CPR revolutionizing writing-to-learn pedagogies, specifically those based on peer-review techniques. However, RHIT faculty has invested a great deal of attention to detail in creating an environment that maximizes CPR's potential. Even more importantly, their reliable and copious data collection will serve to more precisely explain exactly how and why CPR is effective, or not.

It is a bit disappointing that RHIT faculty has not released more results indicating whether CPR is actually improving writing, or learning for that matter. Despite the *in situ* implementation, all of the papers reviewed, with the exception of the ABET assessment analysis, present the results of *in vitro* analyses. In other words, the studies focus more on what is happening inside the CPR module rather than what learning gains are happening (or not) in students.

For example, it would be interesting to conduct a study using the writing assignments archived by CPR. Writing assignments could be holistically assessed for writing competency, with the instructor-evaluator blinded to before and after samples. This kind of comparison could be made across the modules of a single course, or across the portfolio of an entire graduating class.

It is only speculation of course, but perhaps during this early adoption stage RHIT faculty has chosen to focus on adjusting their CPR implementation to determine its greatest beneficial effects. Based on their rigorous application of statistical analyses thus far, one would hope that when they do present data that either validates or invalidates CPR's efficacy, they will also be able to dissect precisely what went right or what went wrong. Certainly the papers presented thus far indicate Carlson and Berry are trying to get as fine-grained an understanding as possible into exactly what is going on inside the software engine.

By fully incorporating CPR into an end-to-end teaching sequence, RHIT faculty has been able to study its functionality in a real environment. As noted in their early papers, Carlson and Berry have been rigorously analyzing data since the days of CPR's initial implementation. Also encouraging is the regular frequency with which they have presented their findings at peer-reviewed conferences. One might describe them as "publishing-to-learn" by sharing and comparing their knowledge with the educational research community. Despite the question of "does it work?" being yet unanswered, these seven papers provide an invaluable archive of breadth and depth.

References

Accreditation Board for Engineering and Technology. (1988-2007). November 2007, <www.abet.org/index.shtml>.

Chapman, O. L. and Fiore, M. A. (2001, February). "The White Paper: A Description of CPR." Los Angeles: University of Los Angeles.

Gerdeman, R.D., Russell, A. A. and Worden, K. J. (2007, March/April). "Web-Based Student Writing and Reviewing in a Large Biology Lecture Course." Journal of College Science Teaching, vol. 36 no. 5, pp. 46-52.

Likkel, L. (2007, January) "Calibrated Peer Review Essays Increase Confidence in Self-assessment." AAS/AAPT Joint Meeting, presentation 170.08. Seattle. (student preference survey)

Margerum, L. D. Gulsrud, M., Manlapez, R., Rebong, R. and Love, A. (2007, February) "Application of Calibrated Peer Review (CPR) Writing Assignments To Enhance Experiments with an Environmental Chemistry Focus." Journal of Chemical Education, vol.

84 no. 2, pp. 292-295. (student satisfaction survey, 10% difference in before/after within same course)

McCarty, T., Parkes, M. V., Anderson, T. T., Mines, J., Skipper, B. J., Grebosky, J. (2005, October) "Improved Patient Notes from Medical Students during Web-Based Teaching Using Faculty-Calibrated Peer Review and Self-Assessment." Academic Medicine Supplement, vol. 80 no. 10, pp. S67-S70. (comparison of before/after med student notes not blinded)

McStravick, D. and O'Mallley, M. (2007, June) "Improving Interdisciplinary Capstone Design Projects With Cooperative Learning In The Medi-Fridge Project." Proceedings of the 2007 American Society For Engineering Education Annual Conference and Exposition, session AC 2007-1674. Honolulu. (Rice University spin-off)

Ohland, M.W., Loughry, M. L., Carter, R. L., Bullard, L.G., Felder, R. M., Finelli, C. J., Layton, R. A. and Schmucker, D. G. (2005, June)"Developing a Peer Evaluation Instrument that is Simple, Reliable, and Valid." Proceedings of the 2005 American Society For Engineering Education Annual Conference and Exposition, session 1526. Portland. (multi-institution spin-off Clemson University, Marymount University, North Carolina State University, University of Michigan, Rose-Hulman Institute of Technology, Western Kentucky University)

Pelaez, N. J. (2002)"Problem-Based Writing with Peer Review Improves Academic Performance In Physiology." Advances in Physiology Education, vol. 26, pp. 174–184. (before (didactic lecture) and after (PW-PR) not blinded)

Plutsky, S. and Wilson, B. A. (2004, Winter). "Comparison of Three Methods for Teaching and Evaluating Writing: A Quasi-experimental Study." The Delta Pi Epsilon journal, vol. XLVI no. I, pp. 50-61.

Prichard, J.R. (2005, Fall) "Writing to Learn: An Evaluation of the Calibrated Peer Review™ Program in Two Neuroscience Courses Psychology Department and Neuroscience Program." The Journal of Undergraduate Neuroscience Education (JUNE), vol. 4 no. 1, pp. A34-A39. (student satisfaction survey)

Russell, A.A. (2004, April). "Calibrated Peer Review™: A Writing and Critical-Thinking Instructional Tool." Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics (STEM) Education. Proceedings of the American Association for the Advancement of Science Conference. Washington, D.C: American Association for the Advancement of Science. pp. 67-70.

Russell, A.A. (2004, June 11). "What Works - A Pedagogy: Calibrated Peer Review." Project Kaleidoscope Volume IV: What works, what matters, what lasts." Washington, D.C.: Project Kaleidoscope. November 2007 < http://www.pkal.org/documents/Vol4CPR.cfm>

Russell, A.A. and Fiore, M.A. (2007, October). CPR Publications (personal communication). Los Angeles: University of Los Angeles.

University of California, Los Angeles. (2001) CPR (Calibrated Peer Review™): Web-based writing and peer review. November 2007 <http://cpr.molsci.ucla.edu/>.

University of California, Los Angeles and University of Southern California. (2005, September). "Calibrated Peer Review: A Writing and Critical-Thinking Instructional Tool." Boulder: EDUCAUSE Learning Initiative.