Calibrated Peer Review: A Tool for Assessing the Process as Well as the Product in Learning Outcomes

Patricia A. Carlson† and Frederick C. Berry††

†Humanities and Social Science Department
††Electrical and Computer Engineering Department
Rose-Hulman Institute of Technology
Terre Haute, Indiana

Abstract: For about two decades now, engineering education has been in the process of re-inventing itself. ABET's revised requirements, changing realities of the workplace, and the growing awareness of language in the learning process all place added emphasis on writing in today's engineering curricula. However, most instructors of engineering design believe themselves to be hard-pressed to incorporate additional writing assignments into courses already filled with content materials. Also, most engineering design instructors may not have either the time or the expertise to provide commentary on student written work. Thus, the formative assessment for these assignments, so critical to learning, doesn't emerge, and the experience may devolve into "busy work" in the eyes of the student. We report on early results from an NSF-funded[1] study using Calibrated Peer Review (CPR) – a web-delivered, collaborative learning environment for writing assignments – in a junior-level introduction to engineering design course. Rather than elaborating on the course itself, we focus in this paper on the data collection capabilities of CPR and give samples of the types of analyses made possible by the tool.

I. Using Writing as an Analog for Thinking in Engineering Design

We base this paper on experiences with Calibrated Peer Review (CPR), largely over the past two years. We demonstrate how the CPR learning platform moves well beyond traditional web-delivered, course enrichment software. More specifically, data collected in a junior-level engineering design course is used to illustrate how CPR system features help to make formative assessment an integral part of using writing in the learning process. In fact, the versatility of the software makes CPR very appropriate as a tool for collecting fine-grained observations appropriate for ABET-style accreditation. We divide our remarks into three segments: (1) a brief overview of the components of CPR, (2) an explanation of the types of data collected by the system, and (3) samples of data analyses to show what types of formative assessment can be provided and how these observations are indicative of student learning.

II. What is Calibrated Peer Review?

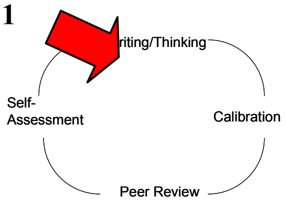
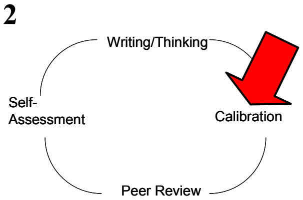
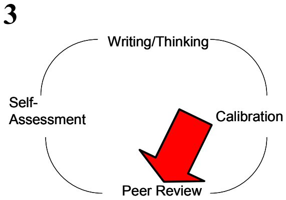
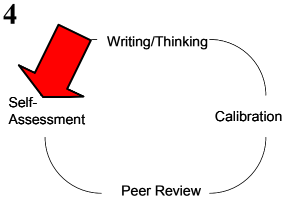
Developed by a group at UCLA led by Dr. Orville Chapman and Dr. Arlene Russell, Calibrated Peer Review is an excellent digital "learning environment" that features an asynchronous, discipline-independent platform for creating, implementing, and evaluating writing assignments,

[1] CCLI - Assessment of Student Achievement

without significantly increasing the instructor's workload. The fully integrated CPR software contains an assignment authoring tool for the 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 [1].

Table A explains how CPR's four structured workspaces function as mentoring stages that not only facilitate the accomplishment of a communication task but also help the learner to internalize strategies for later performance of the same or similar tasks, *without the presence of the technology.* These four structured workspaces perform in tandem to create a series of activities that reflect modern pedagogical procedures for using writing in the learning process. Separate instructor and student interfaces provide reports on performance for individual assignments.

TABLE A: FOUR STRUCTURED WORKSPACES OF CPR

SEGMENT	ACTIVITY
1 	**Writing/Thinking (Assignment and Text Entry):** Students are presented with a challenging, relatively short communication task, with guiding questions to act as scaffolding for the demanding cognitive activities involved. Students compose using a word processor, but upload the finished text as an HTML file. Graphics, equations, and scientific notation are supported.
2 	**Calibrations:** After electronic submission of their texts, 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.
3 	**Reviews of Texts Submitted by Classmates**: After becoming a "trained-reader" – and being assigned a credibility weighting – students read and provide written feedback on three (anonymous) peer submissions using the same rubric as the one used in the calibrations. They also assign each submission a holistic score from 1 to 10.
4 	**Self-Assessment:** In this final activity, students evaluate their own submission. As with calibration and peer review, students use the same rubric. Having "trained" on benchmark samples and then applied their newly-gained expertise in evaluating peer texts and graphics, students assess their own submission. Students are encouraged at this time to record comments 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.

Specifically designed to help instructors use writing as a vehicle for concept learning, Calibrated Peer Review moves well beyond many web-delivered courseware products. Comparing CPR to available Course Management Systems (CMS) -- such as Angel™ or WebCT™ -- demonstrates the characteristics of CPR as a sophisticated educational tool. With a CMS, an instructor can create and deliver complex course content, including text, data and data manipulation, graphics, and simulations. Defined student interactions with the content can be tracked and evaluated, either through online quizzes or other, well-specified interactions with the electronic interface. Additionally, students and the instructor can emulate a "community of practice" by sharing thoughts, providing critiques, or sustaining motivation – either through synchronous or asynchronous communication and/or collaboration.

Calibrated Peer Review, on the other hand, 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. David N. Perkins (1986) offers the following definition of a cognitive tool, here defined as a "thinking frame"[2]:

> . . . [A] *representation intended to guide the process of thought, supporting, organizing, and catalyzing that process*. This representation may be verbal, imagistic, even kinesthetic. When well-practiced, it need not be conscious. A thinking frame, in order to organize our thinking, includes information not only about how to proceed but when to proceed in that way (p. 7, italics in the original).

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 mental models that foster expert behaviors by invoking appropriate strategies, conserving and allocating mental energies, and orchestrating steps in staged problem-solving techniques.

While a Course Management System provides for file sharing and communal commenting, no system available today duplicates the following unique features of CPR as a complex, highly orchestrated cognitive tool:

- 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.

As illustrated in Table B, the complete CPR data log captures a number of evaluation items. Students' names are listed alphabetically and numbered in the far left column (in this example, names have been removed for confidentiality). The row associated with each name reports scores on specific segments of the CPR session. At the bottom of the report, class averages are given for each of the twelve categories. (The accompanying key indicates what each of these numbers represents.)

TABLE B: SAMPLE OF INSTRUCTOR'S REPORT FROM A CPR SESSION

51.	99.33	9.33	9.33	100.00	79.17	0.33	10.00	6	0.37	40.00	0.33	40.00
52.	95.34	8.67	8.67	66.67	66.67	1.00	6.67	4	1.57	40.00	0.33	40.00
53.	98.01	8.01	8.01	100.00	66.67	0.00	10.00	6	0.72	40.00	0.01	40.00
54.	90.35	3.68	3.68	100.00	70.83	2.33	6.67	4	1.18	40.00	1.32	40.00
55.	97.01	7.01	7.01	100.00	70.83	1.33	10.00	6	1.15	40.00	0.99	40.00
56.	95.66	8.99	8.99	66.67	70.83	1.67	6.67	4	0.73	40.00	0.99	40.00
Class Averages												
Categories	Overall Grade	Text		Calibrations					Reviews		Self-Assessment	
		Rating	Score	% Style	% Content	Avg. Dev.	Score	RCI	Avg. Dev.	Score	Dev.	Score
Class Averages	87.46	7.83	7.83	87.04	68.13	1.22	7.90	4.87	0.86	37.95	0.96	38.78

KEY TO DATA COLUMNS

Column Category	**Definition**
Overall Grade	Totals from major categories TEXT, CALIBRATION, PEER, and SELF REVIEW; based on 100 points
Text Rating	Holistic Score (1-10); Avg Weighted Score given by 3 classmates
Text Score	Weighted score converted to a percentage of total component points, as set by the instructor
Calibration % Style	Percentage of calibration questions correct in this category
Calibration % Content	Percentage of calibration questions correct in this category
Calibration Avg. Dev.	Average Deviation on scores given for all three benchmark texts
Calibration Score	Style + Content +Retake + Avg Dev = a percentage of the total component points, as set by the instructor
Calibration RCI	Reader Competency Index: Complex Algorithm explained at CPR website http://cpr.molsci.ucla.edu/
Reviews Avg. Dev.	Student's holistic review compared to average of 2 other reviewers. (Summation of 3 reviews)
Reviews Score	Weighted score converted to a percentage of total component points set by instructor
Self-Assessment Deviation	Self-assigned holistic score compared to the average of 3 classmates' ratings

Self-Assessment Score	Weighted score converted to a percentage of total component points set by instructor

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. Viewing both the empirical and the narrative feedback from CPR sessions is very informative for the instructor.

However, in this presentation, we focus only on the utility of a specific subset of the empirical data. As illustrated in Table B, "Overall Grade" gives an assessment of student performance, expressed in traditional terms of a percentage of 100 points. This number is the sum of "Text Score," "Calibration Score," "Reviews Score," and "Self-Assessment Score." The instructor distributes the 100 points across the four workspaces to reflect the degree of importance she has placed on each activity. Data in these five categories are interdependent and are summative in nature. While important, they are not central to the formative assessment conducted in our pilot study.

On the other hand, data from the other seven categories measure dimensions of the process of learning through writing for individual students or groups of students. Table C explains what these seven measurements are, how they are represented, and why they are useful as formative feedback.

TABLE C: 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.

III. CPR and the Assessment of Learning through Writing

CPR -- as an advanced form of educational technology -- partners both with the student and with the instructor to foster learning. In this project, through the vehicle of CPR, we were able to implement assignments that fully utilize the "writing across the curriculum" (WAC) pedagogy [3 - 6], without overly increasing the workload for instructors. Furthermore, CPR's ability both to elicit and to report qualitative and quantitative peer review helps to make formative assessment an integral part of instruction. And, the data collected by CPR during the student's process of engaging the assignment gives faculty a deeper understanding of how students learn through writing assignments, resulting in better, more individualized feedback for learners. CPR's extensive data summaries also allow for analysis of patterns and trends in aggregates of students, resulting in better faculty awareness in designing instructional materials for maximal benefit.

Our purpose in this pilot study was to explore the types of statistical analyses available to an instructor using CPR, and the pedagogical insights these analyses might provide. We used a junior-level, introductory course in engineering design as our test bed. Ideally, all courses in the engineering curriculum should address concept-building and problem-solving. However, "design" is the essence of modern-day engineering: "design, above all else, defines the difference between an engineering education and a science education" [7].

In this paper, we do not focus on the content of the course or on the implementation of the CPR assignments for three reasons. First, space limitations make it impossible to cover both the assignments and the data analyses. Second, the CPR system is an "empty engine." Thus, assignments can be unique to an individual instructor. Our focus is to demonstrate what can be done with the CPR data in order to assess student learning within a writing assignment. And third, the basic assignments for the test-bed course are the types of writing typically done in an engineering design course. A more complete description of how CPR was implemented in the course appears in Addendum A.

Offered in the spring quarter (from 5 January to 21 May) of 2004, ECE 362 consisted of two sections, for a total of 55 students (12 women and 43 men). Both sections had the same instructor. In the college bulletin, the course is described as: *System engineering, team project involving conception, design specifications, 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.*

During the 10-week quarter, all students participated in six *major* CPR sessions. Table C gives additional information on these writing assignments. Each is relatively standard in engineering design courses.

TABLE D: ASSIGNMENTS CONTAINED IN SIX CPR SESSIONS

Assignment Title	**CPR #**	**Description**	**Dates**
What is a Patent? Assignment #1	Session I	Annotated bibliographic entries for three patents on devices similar to the student's proposed project	Start: 3/12/04 Finish: 3/18/04
Market Analysis Assignment #2	Session II	Based on research, indicate target demographics, needs assessment, market dynamics, and major competitors	Start: 3/19/04 Finish: 3/29/04
Product Design Specification (PDS) Assignment #3	Session III	A definitive statement to the design team giving the requirements and constraints that the new design must fulfill.	Start: 4/11/04 Finish: 4/15/04
Product Technical Description Assignment #4	Session IV	Distinct from the PDS, this document describes salient physical characteristics of appearance and composition.	Start: 4/22/04 Finish: 4/30/04
Social Impact Statement Assignment #5	Session V	One or two paragraphs; considers the larger context / ramifications, which may include political, social, ethical, legal, professional, or physical dimensions.	Start: 5/07/04 Finish: 5/13/04
Product Development Proposal Assignment #6	Session VI	Consolidation: Upload the 5-page project proposal representing the culmination of previous assignments.	Start: 5/18/04 Finish: 5/20/04

This paper describes the analytical methods made possible by the data collected *in-situ* by CPR. However, any instructor considering CPR for course adoption will first want to know about such pragmatic issues as ease of use, return on time investment, and student reaction and learning gains.

- Investment of Course Time: Each instructor decides how much emphasis can (or should) be given to CPR assignments in a given course. Two 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.

- 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.

- Student Reactions: 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. Because each CPR assignment highlights a critical component of the larger, final proposal, 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.[2] 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 nuances they have learned come from scrutinizing the submissions of fellow students. Also, a simple comparison between the final proposals from the pilot course and final proposals from the same course as taught in prior years (without CPR) shows dramatic improvement.

In the following examples, each illustration presents a hypothesis about a specific aspect of student performance within a CPR session or series of sessions from our test-bed course. However, these samples are not intended as a unified study. Rather, they demonstrate how each of the seven data points discussed above can be viewed as variables and analyzed to understand – at a deeper level – the dynamics of learning through writing. This understanding can – in turn – be used by the instructor to council students better on how to improve their performance. The data results might also be used for ABET-style program assessment.

Our target audience is engineering faculty who feel a responsibility to integrate meaningful writing assignments into technical courses, but who are unsure how best to do this. Recent journal articles and conference papers provide accounts of solid programs where communication faculty partner with technical faculty to create rich and challenging courses. However, we posit that these collaborations are the exception rather than the rule. Specifically, we address faculty who don't have access to showcase programs, who may be feeling pressured to include writing without being given the resources to do so, and who may be feeling reluctant, dubious, or even anxious about these new expectations.

---

[2] In a parallel study, we are examining the results from the pilot course to determine if CPR fosters cognitive development as defined by the six stages in Bloom's Taxonomy of Cognitive Objectives.

The following series of "mini-studies" demonstrates four specific types of statistical calculations we did using CPR data: (1) descriptive analysis, (2) correlation of means, (3) calculations of variances among variables, and (4) analysis of a sequence of assignments. We used SPSS to perform these sample analyses; however, the calculations can also be done quite easily in any statistical software package, including *Microsoft Excel* (or other packages commonly available to engineering faculty).

While the statistical results are interesting in themselves, our more compelling purpose is to persuade the target audience that CPR provides a rich learning experience through writing, without overtaxing the instructor. CPR's bases its pedagogy on a model involving sequenced learning episodes, primary of which are explicit training followed by application to peer-produced samples. The five mini-studies presented here deconstruct the sequence and demonstrate the fundamental soundness of the pedagogy instantiated in CPR as a learning tool.

A. *Descriptive Analysis*

Case #1: Looking at measures of central tendency, frequency distribution, standard deviation, and the like summarizes CPR session data in a clear and understandable way. Our first example considers the hypothesis that students who do well in the training portion of the assignment (that is, in the calibration workspace) should be able to apply this training to new but similar situations. (The CPR data point indicating ability to apply learning from the calibration workspace is the average deviation calculated in the Text Review workspace.)

Figure 1 shows a scatter diagram of the variable "revdev" (deviation in review averages) plotted against the variable "reader competency index" (a measure of performance in the calibration). A regression line was fitted to these occurrences. The assignment being plotted is the Product Technical Description, or CPR Session #4, where the number of students completing the task and included in the sample was 47. "Revdev" is the average deviation between a student's holistic score of three peer submissions and the holistic ratings given by the other raters of the same submission. Thus, "revdev" is a measure of how well the student can apply the criteria taught in the calibration section. "RCI" (Reader Competency Index) is a number – from 1 (lowest) to 6 (highest) – built from several measures of performance during the calibration workspace. Thus, one can predict that a high "RCI" should result in improved performance during the peer-review workspace.

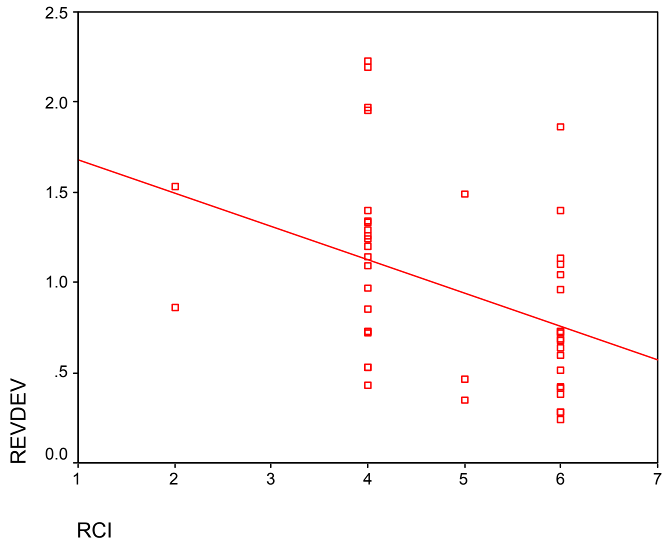


Figure 1: Scatter Plot of Two Variables from Assignment #4

As can be seen, the prediction is supported: students who have performed well during the training phase of CPR Session #4 are also able to apply that learning to a set of student-authored Product Technical Descriptions. The instructor then has strong evidence that the writing assignment is working well and that learning had taken place. *As the data indicate, the students who first learn to recognize the rhetorical features of a good technical description can then apply these same criteria to writing submitted by their classroom colleagues. We believe these results support claims for the efficacy of CPR as a teaching tool.*

Case #2: A second illustration of descriptive analysis further demonstrates how we used CPR data to examine specific episodes in the learning process within the context of a writing-enhanced course. In this case, we looked at all six of the CPR sessions (as described in Table D), and we plotted the mean of the variable "revdev" (deviation in review scores) for each session.

Adding to our analysis a dimension of how students' initial aptitude affects performance, we did a median split on all six sessions, using the variable "text rate" as an indicator of the students' pre-training understanding of the assignment.

Keep in mind the sequence of events within the Calibrated Peer Review environment. The student reads an assignment, produces a text artifact, and then is trained on what an "expert" would look for in evaluating a response to the assignment. After the training, the student then applies this learning to a review of peer-written submissions (variable "revdev"), and then to her own submission (variable "sadev"). Thus, how well a student's submission is rated becomes a proxy-measure of how well the student performed *before* the training in the calibration workspace.

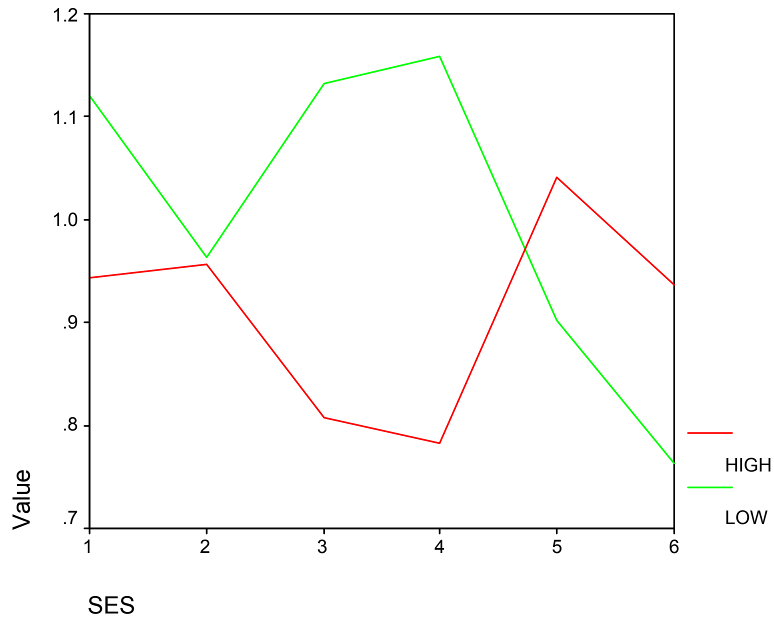


Figure 2: Average Deviation in Peer Review Plotted Across Six Sessions for Median Split Using Text Rate Score

Because "revdev" is a measure of average deviation, the smaller the number, the better the student has performed. As can be seen in Figure 2, these preliminary results do not indicate a clear and consistent negative slope for either ability group.

Nevertheless, we found three observations of interest. First, of all the assignments, the writing prompt in assignment #2 seems to dampen whatever effect initial competency had between the two aptitude groups. Second, assignment #4 clearly treated the two aptitude groups very differently. 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 phase of assignment #4. Third, the training in assignment #6 appears to have worked well for both groups, but seems to have been especially effective with the lower aptitude group. We are examining the 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 using the CPR environment.

The two samples given above are descriptive only. In order to determine whether any of the relationships are statistically significant, or to account for the variance, or to prove causal or predictive relationships among the variables, we needed to do a different level of statistical analysis.

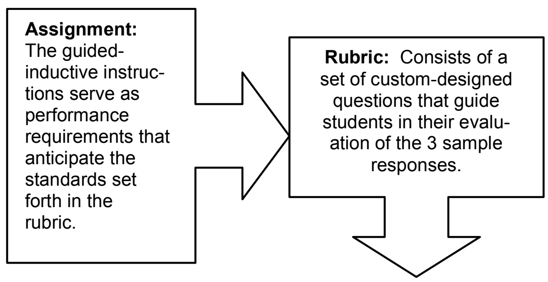
B. *Measures of Correlation*

<u>Case #3:</u> Calculating correlations and determining their significance proved useful in diagnosing student learning. The sample presented here focuses on the more fine-grained relationships

occurring within the variables tracked in the calibration workspace. Our hypothesis is that students who do well on the questions asked in the training rubric (presented as two numbers based on a percentage of 100) should then be able to evaluate samples holistically – that is, apply the concepts of the rubric items – with less error.

The results of the calibration workspace are reported in five numbers, as indicated in Figure 3. (Note that "Score" is not a separate measure; rather, it is a composite result derived by CPR's algorithms and used as an element in determining the student's overall "grade" for performance in all four CPR workspaces. As such, this number is not of interest in the following analysis.)

Both the "% Style" and "% Content" column indicate the percentage of questions about the calibration texts for which the student's answer matched that of the "answer" indicated by the instructor.



Student	%Style	%Content	Avg Dev	Score	RCI
SMITH	55.55	45.45	1.33	1.67	3
JONES	88.89	100.00	0.00	10.00	5

Figure 3: Indicators Tracked in the Calibration Workspace

Each of three benchmark texts is also given a holistic rating from 1 to 10 by the assignment author. The "average deviation" (avg. dev.) number indicates how far away from the rating standard the student was for all three of the calibration texts. As in previous examples, the "RCI" (Reviewer Competency Index) – an integer from 1 to 6 – indicates how well the student "trained" during the calibration. This number is used later in the session – especially in the peer review workspace – to determine a "weighting" for this student's rating of the three peer texts she reviews. (The algorithms embedded in CPR are beyond the scope of this presentation. Suffice it to say that the calculations are very robust.)

A simple case demonstrates how we used the data as a formative assessment indicator. A strong *negative* correlation between either "style" or "content" (or their average) and "avg. dev." indicates that the activity has taught the student well. In other words, the higher the percentage of questions the student can answer correctly either in "style" or in "content" (or both) the better she should be able to apply the rubric and lower her deviation from the instructor-determined norm on the holistic number for the sample text. In other words, as one number increases, the other should decrease, if learning is taking place.

Table E presents correlation coefficients and their significance for variables “style,” “content,” and “caldev” from assignment #3, “Product Design Specification.” The 46 students who completed the task constitute the sample. Notice that both “style” and “content” have a significant, positive correlation (at the 95 % confidence level). This correlation is reasonably strong and is what one would predict for students who perform well in the rubric segment of calibration. Also indicating that learning is taking place, the relatively strong, but negative correlation between both style and content and the average deviation (indicated in “caldev”) is significant at the 99% confidence level.

From this analysis, we determined that assignment #3 was working well, and that students were able to apply the expertise they were learning in the calibration workspace. In short, students who could answer correctly specific questions about the rhetorical features of a product design specification (that is, perform an analytical evaluation) could also use these observations as a basis for correctly judging the overall quality of the document (that is, perform a holistic evaluation). Helping student to develop mature judgment in engineering practice is difficult; capturing evidence that such maturation is taking place is even more difficult. We believe the inferences that can be drawn for CPR indicate that higher-order learning blossoms within the context of this online cognitive tool.

Table E: Coefficients among Style, Content, and Caldev for Assignment #4

Correlations

		STYLE	CONTENT	CALDEV
STYLE	Pearson Correlation	1	.340*	-.487**
	Sig. (1-tailed)	.	.010	.000
	N	46	46	46
CONTENT	Pearson Correlation	.340*	1	-.444**
	Sig. (1-tailed)	.010	.	.001
	N	46	46	46
CALDEV	Pearson Correlation	-.487**	-.444**	1
	Sig. (1-tailed)	.000	.001	.
	N	46	46	46

*. Correlation is significant at the 0.05 level (1-tailed).

**. Correlation is significant at the 0.01 level (1-tailed).

<u>Case #4:</u> A second illustration of how we used correlation coefficients adds an “initial aptitude” variable. Using assignment #6 (Product Development Proposal) and splitting the sample of 47 students (based on a median sort using SAT Verbal score) yields two groups, a nominally “high” aptitude (n= 22) and a nominally “lower” aptitude (n= 27).

The hypothesis being tested is that results from training (indicated by “caldev”) and ability to apply the concepts from the training (indicated by “revdev”) should be positively correlated if learning is taking place. As indicated in Tables F and G, analyzing the “lower” aptitude group of 22 students produced no significant results. Furthermore, the weak correlation between caldev and revdev is negative, a puzzling result.

However, looking at the "high" SAT Verbal group indicates a strong positive relationship between the variables, one that is significant at the 99% level. From these tables, one can infer that the calibration segment of assignment #6 demonstrates an aptitude-treatment effect. Lower ability students did not improve their performance after training; however, the higher ability students did. This gave us cause to re-examine the calibration component of assignment #6.

TABLE F: COEFFICIENTS BETWEEN CALDEV AND REVDEV FOR ASSIGNMENT #6, LOWER ABILITY GROUP

Correlations[a]

		CALDEV6	REVDEV6
CALDEV6	Pearson Correlation	1	-.196
	Sig. (1-tailed)	.	.191
	N	22	22
REVDEV6	Pearson Correlation	-.196	1
	Sig. (1-tailed)	.191	.
	N	22	22

a. NTILES of SATV = 1

TABLE G: COEFFICIENTS BETWEEN CALDEV AND REVDEV FOR ASSIGNMENT #6, HIGH ABILITY GROUP

Correlations[a]

		CALDEV6	REVDEV6
CALDEV6	Pearson Correlation	1	.571**
	Sig. (1-tailed)	.	.001
	N	27	27
REVDEV6	Pearson Correlation	.571**	1
	Sig. (1-tailed)	.001	.
	N	27	27

**. Correlation is significant at the 0.01 level (1-tailed).

a. NTILES of SATV = 2

C. *Analyses across Assignments*

The CPR functionality described suggests that CPR is – *de facto* – a form of electronic portfolio. Certainly, the basic characteristics of (a) student account structures, (b) durable storage of artifacts, (c) course and academic life-span management tools, and (d) evaluative information, would permit a portfolio-like implementation. The CPR system administrator creates an account for an individual student, and the student's name is added to each course using CPR; files tagged as from specified courses are maintained in the student's record. Over the student's learning career, a collection of CPR results builds up – potentially from freshman to senior courses. Results exported to a spreadsheet become a "database" that may be addressed in a number of different ways, giving multiple "views" on an individual or a corpus of students, assignments, or clusters of assignments.

It is fair to say that the majority of portfolio assessments (electronic or otherwise) as they are now used in engineering education are *deductive* assessment tools. The student is given a set of educational goals, and asked to use her judgment in selecting and explaining why certain artifacts validate learning across one or more categories. CPR – on the other hand – acts as an *inductive* assessment tool. The system captures specific observations during learning episodes and the evaluation accrues from feature analyses of written artifacts and learning activities. Patterns may emerge in this data, from which general conclusions about performance can be advanced. Mostly importantly, the portfolio monitors both the processes and the products of learning.

Case #5: As our final example, we demonstrate how a composite of data collected from each of a series of assignments can be proven – through inferential statistics – to have predictive value. The method employed here is regression analysis, a technique used to find relationships between variables for the purpose of predicting future values.

The hypothesis being tested is that the variable "text rating" (the holistic evaluation provided by peer raters) serves as a measure of "pre-treatment" aptitude and has predictive power for student performance (learning) in the calibration workspace. More specifically, the regression reported in Table H represents a test of the supposition that high text ratings should result in a low deviation in the application segment of calibration (that is, a low number in the variable "caldev"). As can be seen, a weak – but highly significant – inverse relationship was found between the variable representing aptitude and the variable representing learning. In short, one can cautiously say that the students' who enter the CPR assignment with a better understanding of the task are able to ingest the training and apply that knowledge better than those who do not. On the other hand, one can also suggest from this data, that for this segment of six assignments, the pedagogy of CPR did not favor those who entered the system better prepared to perform the required writing tasks.

Table G: The Impact of Aptitude and Learning on the Ability to Critically Review Text: Regression Results from Combining Results from All Homework Sessions

Dependent Variable: Deviation from others in critically reviewing text	
Independent Variables	
Intercept	1.188*** (6.331)
Text rating: Peer rating of student text (aptitude)	-0.059*** (-2.409)
Calibration: Incorrect answers (learning)	0.155*** (3.677)
Other Regression Statistics	
R-squared	0.064
F-test	9.139†††
Number of observations	269

t-statistics are in parentheses. *** indicates the null hypothesis that the coefficient is either zero or the opposite sign is rejected at the 1% level, using a one-tail, statistical hypothesis test. ††† indicates the null hypothesis that both the slope coefficients associated with aptitude and learning are simultaneously equal to zero is rejected at the 1% level.

Conclusion

The data used in this paper are from a pilot project that used writing/communication assignments to improve the teaching of engineering design. We based our approach on the "writing across the curriculum" (WAC) movement's premise that verbal composition is an analog for thinking and that communication artifacts can be used to infer student learning in complex problem-solving situations [8 – 13]. Enacting the notion that writing is an analog for thinking – we believe – enables richer, more effective methods of instruction in design – an area of engineering education requiring students to develop mature thought processes in a relatively brief period. Specifically, we believe that closely examining both the *product* and the *process* of writing gives an instructor insights into students' understanding (as well as misconceptions) in ways potentially more profound than assessing other learning outcome artifacts. Furthermore, having such a "window" on a student's thought processes gives an instructor many more opportunities to provide commentary and guidance, the formative assessment so critical to improved future performance. Figure 4 illustrates how CPR instantiates a continuous feedback loop within a course.

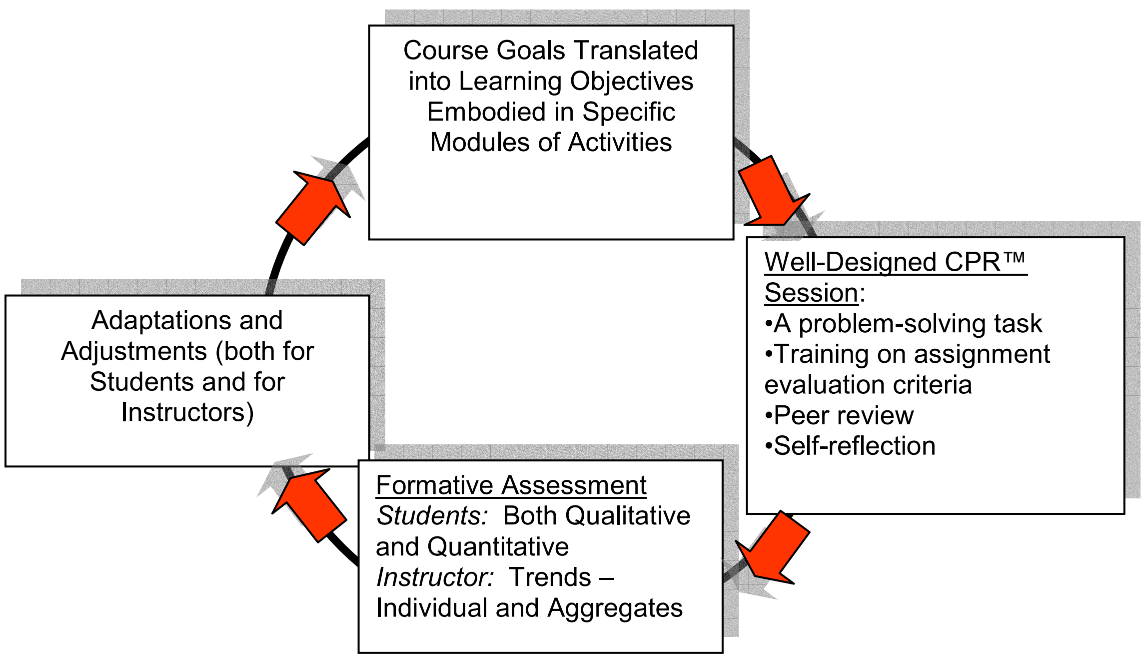


Figure 4: Continuous Feedback Loop for Course Design and Enactment

The context for this paper is the premise that moving students toward mature professional practice requires better methods for teaching engineering design. This paper has implicitly argued that Calibrated Peer Review embodies four concepts central to the current dialogue on "reinventing" engineering education:

1. A innovative pedagogical method (using writing as an heuristic for learning complex thinking and mature behavioral skills);

2. An advanced socio-cognitive, computer-mediated learning tool (CPR);
3. A challenging area of engineering education (teaching design);
4. The pragmatics of program assessment for ABET-style accreditation.

Acknowledgement

This material is based upon work supported by the National Science Foundation (NSF) under Grant Numbers 9980867 and 0404923.

.

References

[1] Calibrated Peer Review™ Webpage. [Online]. Available http://cpr.molsci.ucla.edu

[2] Perkins, D. N. “Thinking frames,” *Educational Leadership*, vol. 43, pp. 4-10, 1986.

[3] Emig, J. “Writing as a mode of learning,” *College Composition and Communication*, vol. 28, pp. 122-128, 1977.

[4] Mcmillen, L., “Science and math professors are assigning writing drills to focus students’ thinking,” *The Chronicle of Higher Education*, Jan. 22, 1986.

[5] Driskill, L. “Students’ reasoning and rhetorical knowledge in first-year chemistry,” *Language and Learning Across the Disciplines*, vol. 2, no. 3, pp. 3-24, 1998.

[6] Moore, R., “Does writing about science improve learning about science?,” *Journal of College Science Teaching*, pp. 95-98, November 1982.

[7] Hodge, B.K. and Steele, W.G., “Experiences with a curriculum with balanced design content in all stems,” *Proceedings*, ASEE Annual Conference, vol. 1, pp. 225-211, 1995.

[8] Driskill, L. “Linking industry best practices and EC3(g) assessment in engineering communication,” *Proceedings,* American Society for Engineering Education Conference*,* 2000. [Online]. Available: http://www.ruf.rice.edu/~engicomm/public/Driskill.html

[9] Gruber, S, Larson, D, Neville, M., and Scott, D., "Writing 4 practice in engineering courses: Implementation and assessment approaches." *Technical Communication Quarterly*, vol. 8, no. 4, pp. 419-441, 1999.

[10] DeWolf, J. T., “Incorporating of writing into a steel design course,” *Journal of Professional Issues in Engineering Education & Practice*, vol. 128, pp. 71-75, 2002.

[11] Pytlik, B., “Sequencing writing assignments to foster critical thinking,” T. Capossela (Ed.), *The Critical Writing Workshop*, Portsmouth, NH: Boynton/Cook, 1993.

[12] Swarts, J. and Odell, L., “Rethinking the evaluation of writing in engineering courses,” *Proceedings,* ASEE/IEEE Frontiers in Education Conference*,* 2001*,* pp. TCA-25 – 30. [Online]. Available http://www.ruf.rice.edu/~engicomm/public/SwartsOdellFIE.PDF

[13] Chenoweth, N.A., Hayes, J.R., Gripp, P., Littleton, E. B., Steinberg, E.R., and Van Every, D. “Are our courses working? Measuring student learning,” *Written Communication*, vol. 16, pp. 29-50, 1999.

Biographies

PATRICIA A. CARLSON is a long-time advocate of writing in engineering education. She has been a National Research Council Senior Fellow for the U. S. Air Forcer, as well as having had several research fellowships with NASA (Langley and Goddard) and the Army’s Aberdeen Proving Ground. She has also been a research fellow at NASA’s Classroom of the Future located in Wheeling, WVA. Her primary research area – computer-aided tools to enhance writing in engineering education – has been funded through two NSF grants.

FREDERICK C. BERRY is the Head of the Electrical and Computer Engineering Department at Rose-Hulman Institute. Long an advocate of engineering educational reform, he has been active in such innovations as project learning, entrepreneurial studies, and advanced IT in the classroom. He also participates in many outreach programs to interest middle school children in engineering.

Addendum A

HOW WE APPLIED CPR

Most engineering programs have some type of capstone design experience. At Rose-Hulman Institute of Technology the Electrical and Computer Engineering (ECE) Department as taken this a step farther by having a course (Engineering Practice) which teaches the fundamentals of design before the students in ECE start their capstone design experience.

This course, Engineering Practice, is for the most part a technical writing course. The concepts of research, project design specification, high level design, detailed design, work break down schedules, budgets, and teaming are taught to the students by having them write a proposal for their senior design project.

THE FIRST 3 WEEKS

In the first 3 weeks of the 10 week term the students do research and idea generation.

CPR 1: What Is Intellectual Property (IP)

This CPR introduces IP in the form of patents, trademarks, industrial designs and copyright to the students. Patent protection is the major focus of this CPR. Students review several patents and learn an idea must be**:** novel, non-obvious, or useful. Then the students try to generate three ideas each that are novel, non-obvious, or useful.

CPR 2: What Is an Annotated Bibliography?

This CPR introduces students to research using the annotated bibliography. The reason the annotated bibliography is used, it adds descriptive and evaluative comments (i.e., an annotation), assessing the nature and value of the cited works. The addition of commentary provides critical information and a foundation for further research. Therefore, the students are required to read and assess the value of a reference to the project they are researching. The students are required to generate five annotations for their project. With a team of three-four students this generates 15-20 annotations. In most cases the students are researching several ideas and have 8-10 annotations per idea per student.

CPR 3: Market Analysis

The students are introduced to two methods of market analysis coupled with project idea generation. These two methods are augment or bi-associate projects.

Augmented Projects are existing products that are *added-to or supplemented*, to extend their functionality. These types of projects are the easiest to do since the base product is already developed. It is also easy to get market information on these types of products.

Bi-associated Projects are projects that *combine two different* products and create a new product from the combination. These types of projects are more difficult to do since the combination of technologies or products may not be so obvious. However, it is still easy to obtain market information for each product and then estimate a market if the two different products were combined into one product.

Each student is required to submit three products that are augmented or bi-associated. About 75% of the project ideas are augmented projects.

THE NEXT 6 WEEKS

In the next 6 weeks the students use the research and experience gained from the first 3 weeks to start writing their proposal for senior design.

CPR 4: Product Design Specification

A Project Design Specification (PDS) is a document that will change substantially over the length of the project. There are many factors that will cause a PDS to change. But the one factor that will have the greatest impact is the development of a deeper understanding of the project. As the student teams move forward developing their project proposal, they will always need to think more intensely about their project idea. The PDS should reflect the common knowledge of the team about the project idea. Therefore, the PDS needs to be regularly refined during the proposal phase to reflect a deeper understanding of the team's project idea.

To develop this deeper understanding the teams use their initial research to develop a HLDD (high level design document). From this research, performance and environmental specifications are developed for their project. These performance and environmental specifications need to accurately reflect the team's present understanding of the project and to be as specific as possible about what the project will do.

Performance specifications address a need. For examples, a pipe positioning system must deliver accuracies of +/- 5 inches for each linear mile of pipe placed.

Environmental specification addresses the surroundings and conditions of operation. For example, the pipe positioning system will use trenchless technology to place pipe underground.

The student teams produce a preliminary PDS at this point for their project.

CPR 5: Project Technical Description

The project technical description should provide a concise explanation, which is not overly technical while frequently emphasizing the key benefits and incorporating appropriate visual elements. Therefore the three essential elements of the project technical description are:

1. Description: It is important to start the description with a very concise description in order to put the features and benefits in context.

2. Visual Element: A picture, a sketch, screen shot, or a diagram that shows either the components of the product or how the product fits in its environment is usually helpful for the reader.
3. Key Benefits: State the key benefits of the product early. The use of bullet points is ideal. Then conclude stating the key benefits again in a paragraph form.

The students produce their first draft of the project technical description using the information from the previous CPRs.

CPR 6: Project Technical Description

The students next take the feedback from CPR 5 and rewrite their project technical description with these specific elements:

1. Does the project technical description tell the reader what the product does in the opening paragraph or sentence?
2. Does the project technical description use concise and precise sentences along with concrete words to explain the product?
3. Does the project technical description use visual elements to help explain the product?
4. Does the project technical description present the key benefits of the product early in the description?
5. Does the project technical description present an analysis of any competitors?
6. Does the project technical description include an explanation of how the parts fit and function together?
7. Does the project technical description conclude with the key benefits of the product in paragraph form near the end of the description?
8. Does the project technical description convince you this project can be done?

The students are also using the NCIIA E-Team RFP as a format guide for the project technical description.

CPR 7: Product Design Specification, Again

The teams have been refining their PDS as they move forward developing their project proposals. Therefore the PDS is reviewed again using the following questions:

1. Is a function list given with a short description for each project-function?
2. Are performance specification given for each function?
3. Is the operating environment for the project given?
4. Are specifications provided relating to the operating environment provided?
5. Are target technologies identified to meet all of above?

At this point the PDS for each student team is very well structured.

CPR 8: Social Impact Statement

This CPR requires the students to analyze their proposed project and write a social impact statement using the IEEE Code of Ethics as the rubric. For this assignment the students write one or two paragraphs about the impact of their project on society.

Other CPRs

Additional CPRs are done in this course: resume construction, memo writing, writing an executive summary, and how to do power point presentation.