International Journal for the Scholarship of Teaching and Learning

Volume 3 | Number 2 Article 15

7-2009

Improvement in Writing and Reviewing Skills with Calibrated Peer ReviewTM

Adalet Baris Gunersel
Philadelphia, Pennsylvania, USA, bgunersel@temple.edu

Nancy Simpson
College Station, Texas, USA, n-simpson@tamu.edu

Recommended Citation

Gunersel, Adalet Baris and Simpson, Nancy (2009) "Improvement in Writing and Reviewing Skills with Calibrated Peer ReviewTM," *International Journal for the Scholarship of Teaching and Learning*: Vol. 3: No. 2, Article 15.
Available at: https://doi.org/10.20429/ijsotl.2009.030215

Improvement in Writing and Reviewing Skills with Calibrated Peer ReviewTM

Abstract

This study analyzed results of an NSF-funded project that used Calibrated Peer Review (CPR)TM to promote writing and reviewing skills. The specific focus of the study was whether students at different levels of performance showed improvement in writing and reviewing competency with repeated use of CPR. The study paid specific attention to progress made by initially lower performing students. The courses of nine instructors with a total of 789 students were included. Repeated measures analyses indicated that across different science disciplines and student levels, students showed improvement in writing skills and reviewer competency with repeated use of CPR. In addition, the difference in scores between high and low performing students decreased over time in both writing skills and reviewer competency.

Keywords

Calibrated Peer Review, Writing skills, Critical thinking skills, Undergraduate science education

Improvement in Writing and Reviewing Skills with Calibrated Peer Review™

Adalet Baris Gunersel
Temple University
Philadelphia, Pennsylvania, USA
bgunersel@temple.edu

Nancy Simpson Texas
A&M University College
Station, Texas, USA n-
simpson@tamu.edu

Abstract

This study analyzed results of an NSF-funded project that used Calibrated Peer Review (CPR)™ to promote writing and reviewing skills. The specific focus of the study was whether students at different levels of performance showed improvement in writing and reviewing competency with repeated use of CPR. The study paid specific attention to progress made by initially lower performing students. The courses of nine instructors with a total of 789 students were included. Repeated measures analyses indicated that across different science disciplines and student levels, students showed improvement in writing skills and reviewer competency with repeated use of CPR. In addition, the difference in scores between high and low performing students decreased over time in both writing skills and reviewer competency.

Keywords: Calibrated Peer Review, writing skills, critical thinking skills, undergraduate science education

Introduction

This study focused on the courses of instructors who participated in the Writing for Assessment and Learning in the Natural and Mathematical Sciences (WALS) Project, funded by the National Science Foundation. The project adapted an innovative teaching tool, Calibrated Peer Review (CPR) ™, in Biology, Physics, and Mathematics at a large land-grant university to better assess student understanding, to enhance student learning, and to observe the integration of writing into these science courses (http://cpr.molsci.ucla.edu/). Earlier studies demonstrated that students who received low scores on initial CPR assignments showed progress throughout the semester, indicating improvement in their writing and reviewing skills (Gerdeman, Russell, & Worden, 2007; Gunersel, Simpson, Aufderheide, & Wang, 2008). The current study investigated whether this pattern holds for the students of the instructors participating in the WALS Project. The research question of the study was: Did initially lower-performing students show improvement in writing and reviewing competency with repeated use of CPR? The study involved data from courses of nine faculty members, each of whom used at least three CPR assignments; a total of 789 students were included.

https://doi.org/10.20429/ijsotl.2009.030215

Calibrated Peer Review

Developed at UCLA for the Molecular Science Project, one of the NSF-supported Chemistry Systemic Reform Initiatives, CPR was designed to give students practice in writing and peer review, since both are expected competencies in scientific fields (Russell, 2001). One of CPR's aims is to develop students' skills of discipline-specific writing, a prominent educational goal (Emerson, MacKay, MacKay, & Funnell, 2006; Lea & Street, 1998). The underlying pedagogy of CPR is supported by numerous studies demonstrating the educational value of both writing (Holliday, Yore, & Alvermann, 1994; Klein, 1999; Kovac & Sherwood, 1999; Lowman, 1996; Rivard, Stanley, & Straw, 2000) and peer review (Falchikov, 1995; Orsmond, Merry, & Callaghan, 2004; Searby & Ewers, 1997; Sluijsmans, Brand-Gruwel, & Van Merrienboer, 2002; Sluijmans, Docky, & Moerkerke, 1999; Topping, 1998), which are desired skills searched for by employers.

To increase the ability of students to review their peers' work, CPR includes a "calibration phase" during which students practice reviewing according to the instructor-designed rubric.

In order to create a CPR assignment, instructors produce the following components:

Instructions for writing.
Instructions include suggested resources, questions to guide student thinking, and a "writing prompt" that tells students such things as the topic, format and audience for their writing.

Calibration questions.
Calibration questions direct students' attention to content and style characteristics of a completed assignment and form the basis for assigning a text rating.

Three sample essays.
The high, average, and low quality sample essays are responses to the assignment and that have been evaluated by the instructor using the calibration questions.

Student work on a CPR assignment occurs in three phases:

Text entry phase
Students read instructions, access suggested resources, and write and submit their essays.

Calibration phase
Students are presented with the three sample essays, along with the calibration questions. For each essay, students answer the calibration questions and assign a rating. CPR assigns a reviewer competency index based on a comparison of the student review to the instructor review of each essay.

Review phase
Students are presented first with three classmates' essays (randomly assigned and anonymous) and then with their own essay, all of which they review and rate using the same set of calibration questions.

Instructor-reported experiences and previous studies indicate that CPR is a tool that can help students master content, improve writing skills, and become more competent reviewers (Furman & Robinson, 2003; Gunersel, Simpson, Aufderheide, & Wang, 2008; Hand, Hohenshell, & Prain, 2007; Hartberg, Gunersel, Simpson, & Balester, 2008; Keeney-Kennicutt, Gunersel, & Simpson, 2008; McCarty, Parkes, Anderson, Mines, Skipper, & Greboksy, 2005; Russell, 2001). Previous studies used CPR-generated scores that measure writing and reviewing skills to investigate whether students improved in such areas. For example, Gerdeman, Russell, and Worden (2007) found that CPR-generated scores of 1330 students in an introductory biology course showed statistically significant increases, suggesting that their writing and reviewing abilities also improved. In addition to this, they found that students whose scores were initially lower than the others' showed the greatest improvement. Although this could be the result of "regression to the mean" which suggests that initially low scores would be more likely to increase, the authors concluded that it was the result of CPR's effect. Gunersel, Simpson, Aufderheide, and Wang (2008) found that repeated use of CPR improved the writing skills of 47 students and the reviewing skills of 84 students in a senior-level biology course. In another study which included a Likert scale survey, more than 50% of first-semester general chemistry students "agreed" or "strongly agreed" that they were "better technical reviewers" by doing CPR assignments (Margerum, Gulsrud, Manlapez, Rebong, & Love, 2007, p. 294). Pelaez (2002) compared the learning outcomes of 35 undergraduate nonscience majors taught with traditional lectures and taught with CPR in an introductory physiology course. The results indicated that the performance of students who had completed problem-based learning assignments in CPR was better than or equal to the performance of students who had received "traditional instruction" (statistically significant difference at alpha level .01) (p. 181). Pelaez (2002) noted:

> The favorable results may be a product of the work students complete when writing about their thinking, or perhaps students did better because PW-PR (problem-based writing with peer review) made it possible for them to confront and resolve difficulties they encountered relating concepts. (p. 181)

In addition to its benefits for students learning, CPR also has benefits for instructors. A recent study (currently in print) by the authors of this paper found that CPR makes it easier for instructors to use writing assignments in big classes and allows instructors to spend much less time on grading. The one disadvantage of CPR may be the time instructors need to spend on creating effective assignments.

The number of published studies that provide evidence of the value of CPR as a tool for improving students' conceptual learning, writing skills, and critical thinking skills is growing; this study contributes to this body of literature.

Methods

This study investigates this research question: Did initially lower performing students show improvement in writing and reviewing competency with repeated use of CPR in the courses of instructors participating in the WALS Project?

Nine instructors who were a part of the WALS Project between 2003 and 2008 were included in the study. These instructors, who were still a part of the large land-grant university, were selected because they had utilized at least three CPR assignments within

a course. Four instructors were in Biology, two in Mathematics, and three in Physics. Table 1 presents information on disciplines, class levels, numbers of students, semesters, and numbers of assignments. A total of 789 students–only those who had completed all assignments—were included in the study.

In order to investigate students' writing and reviewing competency two CPR-generated scores, reviewer competency index (RCI) and text rating (TR), were used. The RCI is computed (by the CPR program) following student review of three instructor-provided essays. The computation uses a comparison of student and instructor responses to instructor-generated calibration questions, as well as a comparison of student and instructor global ratings of the essays. TR on the other hand, is a weighted average of scores given by three peer reviewers. Weighting is based on reviewing competency (RCI) of the peer reviewer. Reviewers are instructed to base the score on analysis guided by the calibration questions. Since the calibration questions include both content-related questions and writing-related questions, TR can reflect both content understanding and writing competence. In summary, TR is used as a measure of writing quality and content understanding, while RCI is used as a measure of students' ability to review. For each CPR assignment students receive a TR ranging from 1 to 10 and a RCI ranging from 1 to 6.

Table 1. *Information on Participating Instructors*

Instructor Code	***Discipline***	***Class Level***	***# Students***	***Semester/s Included***	***# of Assignments***
A	Biology	300	147	Spring 2004, Spring 2005, Spring 2006, Spring 2007	3
B	Physics	200	74	Fall 2004, Spring 2005, Fall 2005	3
C	Math	200	54	Spring 2004, Fall 2004, Spring 2005	3
D	Biology	400	81	Spring 2005, Spring 2006, Spring 2007	4
E	Physics	200	140	Fall 2005	3
F	Biology	300	48	Fall 2004	3
G	Math	100	52	Fall 2004	5
H	Biology	200	63	Spring 2007	5
I	Physics	200	130	Fall 2004	5

The data was gathered by one of the authors who is an administrator of CPR and thus had access to students' TRs and RCIs within the system. The TRs and RCIs were tabulated in SPSS where the statistical analyses were conducted.

Students were categorized into three groups according to their TRs and RCIs from the first assignment: high (highest 25%), medium (the middle 50%), and low (lowest 25%). In some cases there was not a sufficient range to create three groups and in these cases two groups (higher 50% and lower 50%) were considered. The purpose behind this categorization was to evaluate change in student performance and determine initially lower performing students. Although the study's focus is on initially lower performing students, the progress of the students in the other two categories was also of interest.

Data Analysis

Since the instructors used different CPR assignments in their courses, each course was analyzed separately. Four of the instructors (A, B, C, and D) taught the same course in a few semesters and because the CPR assignments they implemented, as well as the course content and levels of the students, were the same, these semesters were grouped together (Table 1). The separate analyses of the courses also deal with the variability due to different content matters, different fields of the instructors (physics, math, and biology), and different levels of the courses.

Thus, eighteen repeated measures analyses were conducted, two for each instructor. In half of the analyses, the dependent variable was TR, in the other half, the dependent variable was RCI. Performance groups (high, middle, and low) were entered as the grouping variable, while the assignment number was the within-subjects variable. Repeated measures analyses calculated two sets of statistical significance for each instructor's course, presented in Table 2: (a) the change of students' overall TRs and RCIs (presented as "TR(overall)" and "RCI overall" in the table); (b) the change of student performance groups' TRs and RCIs (presented as "Time*TR groups" and "Time*RCI groups" in the table). This study focuses only on the change of the student performance groups, specifically the lower performing group. Since the repeated measures analyses did not indicate which performance group changed, Graph Sets 1, 2, and 3 were created.

Table 2. Repeated Measures Results

Instructor Code		*df*	*F*	*Sig.*	η^2
A					
	TR (overall)	2	4.802	.009	.032
	Time*TR groups	4	15.610	.000	.178
	RCI overall	2	.064	.938	.000
	Time*RCI groups	2	39.296	.000	.213
B					
	TR (overall)	2	4.572	.012	.068
	Time*TR groups	4	7.913	.000	.201
	RCI overall	2	3.931	.022	.052
	Time*RCI groups	2	25.597	.000	.262
C					
	TR (overall)	2	12.590	.000	.198
	Time*TR groups	4	.988	.417	.037
	RCI overall	2	2.652	.075	.049
	Time*RCI groups	4	7.625	.000	.230
D					
	TR (overall)	3	2.943	.034	.042
	Time*TR groups	6	8.601	.000	.204
	RCI overall	3	5.812	.001	.069
	Time*RCI groups	6	23.641	.000	.377
E					
	TR (overall)	2	41.887	.000	.234
	Time*TR groups	4	16.672	.000	.196
	RCI overall	2	11.310	.000	.076
	Time*RCI groups	2	29.313	.000	.175

F					
	TR (overall)	2	2.209	.116	.047
	Time*TR groups	4	4.610	.002	.170
	RCI overall	2	15.401	.000	.251
	Time*RCI groups	2	16.304	.000	.262
G					
	TR (overall)	4	15.868	.000	.245
	Time*TR groups	8	3.036	.003	.110
	RCI overall	4	4.522	.002	.084
	Time*RCI groups	8	2.183	.030	.082
H					
	TR (overall)	4	17.444	.000	.225
	Time*TR groups	8	11.179	.000	.271
	RCI overall	4	20.014	.000	.247
	Time*RCI groups	4	8.830	.000	.126
I					
	TR (overall)	4	6.716	.000	.051
	Time*TR groups	8	9.076	.000	.127
	RCI overall	4	4.456	.002	.034
	Time*RCI groups	8	8.514	.000	118

Results

Results suggest that there was a statistically significant change at alpha level .01 in performance groups' TRs for all of the instructors' courses except for Instructor "C"'s (Table 2). Graph Set 1 shows that TRs of the initially lower performing groups increased from the first assignment to the last in all of the courses, except for Instructor "G"s course, which is presented in Graph Set 3. The significant change in this course demonstrated a different pattern: although the lower performing group's TRs increased from the first to the third assignment, they decreased from the third to the fifth.

In all of the nine instructors' courses, there was a significant change at alpha level .01 or .05 in performance groups' RCIs (Table 2). Graph Set 2 shows that RCIs of the initially lower performing groups significantly increased from the first assignment to the last. Graph Set 3 shows that the change in performance groups' RCIs in Instructor "G"s course demonstrated a different pattern. The lower performing groups RCIs initially increased, decreased, then increased, and decreased again.

Discussion

Previous work indicated that repeated use of CPR facilitates improvement in student writing about scientific topics as well as in their ability to review (Gerdeman, Russell, & Worden, 2007; Gunersel, Simpson, Aufderheide, & Wang, 2008). In particular, these previous studies demonstrated that biology students with low scores (both TR and RCI) on an initial CPR assignment improved significantly on subsequent assignments. This pattern was replicated in almost every case in this current study that involved nine instructors in three different disciplines, with 789 students ranging from first-year college students to graduate

students. The results of the repeated measure analyses presented above reinforce the idea that repeated practice of the type facilitated by CPR is an effective way to help all students—especially those who are initially lower performing—develop their ability to write and review. This study adds to a growing body of literature showing that instructor-guided feedback from peers is able to support this kind of improvement (e.g., Furman and Robinson, 2003; Gerdeman, Russell, & Worden, 2007; Margerum et al., 2007; McCarty et al., 2005; Pelaez, 2002).

While the statistical analyses show that multiple assignments lead to overall improvement in student performance on CPR writing and reviewing, the graphs reveal a more complex picture: the improvement is not monotonic, nor is it uniform for all groups of students. In fact, in some cases, scores of initially high-performing students seem to decrease. This may have been due to students' decreased efforts or "regression to the mean" which suggests that initially high scores would be more likely to decrease. A future qualitative study may investigate why students' scores might show such a trend. Further study is needed to understand why this might be the case. Questions to explore include: Does changing the nature of the assignments diminish the value of repeated practice? Are some learning tasks more suitable for CPR than others? What instructor strategies increase the likelihood that students will give their best effort to CPR assignments? Other studies could include comparison groups, the lack of which is a limitation to the current study. Furthermore, a future mixed-methods study can investigate the reasons behind the statistically significant fluctuations Instructor "G"s course.

Acknowledgement
We wish to acknowledge the consultation and feedback from Dr. Stephanie Knight, Department of Educational Psychology, School Psychology, and Special Education (ESPSE), Penn State University, and Dr. Arlene Russell, Department of Chemistry, University of California, Los Angeles in preparing this manuscript. This material is based upon work supported by the National Science Foundation under Grant No. DUE-0243209.

References

Emerson, L., MacKay, B. R., MacKay, M. B., & Funnell, K. A. (2006). A team of equals: Teaching writing in the sciences. *Educational Action Research, 14*(1), 65-81.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education and Training International, 32*(2), 175-187.

Furman, B., & Robinson, W. (2003). *Improving engineering report writing with Calibrated Peer Review*. Paper presented at the 33rd ASEE/IEEE Frontiers in Education Conference, November 5-8, 2003, Boulder, CO, pp. F3E-14-F3E-15.

Gerdeman, R. D., Russell, A. R., & Worden, K. J. (2007). Web-based student writing and reviewing in a large biology lecture course. *Journal of College Science Teaching* (March/ April 2007), 46-52.

Gunersel, A. B., Simpson, N. J., Aufderheide, K., & Wang, L. (2008). Effectiveness of Calibrated Peer Review™ for improving writing and critical thinking skills in biology undergraduate students. *The Journal of Scholarship of Teaching and Learning,(8)*2, 25-37. http://www.iupui.edu/~josotl/VOL_8/No_2/v8n2gunersel.pdf.

Hand, B., Hohenshell, L., & Prain, V. (2007). Examining the effect of multiple writing tasks on year 10 biology students' understandings of cell and molecular biology concepts. *Instructional Science, 35*(4), 343-373.

Hartberg, Y., Gunersel, A. B., Simpson, N. J., & Balester, V. (2008). Development of Student Writing in Biochemistry Using Calibrated Peer Review*. The Journal of Scholarship of Teaching and Learning, 8*(1). http://www.iupui.edu/~josotl/VOL_8/No_1/V8N1_TOC.htm.

Holliday, W. G., Yore, L. D., & Alvermann, D. E. (1994). The B-science learning-writing connection: Breakthroughs, barriers, and promises. *Journal of Research in Science Teaching, 31*, 877-894.

Keeney-Kennicutt, W. L., Gunersel, A. B., & Simpson, N. J. (2008). Overcoming student resistance to a teaching innovation. *The International Journal for the Scholarship of Teaching and Learning, 2*(1). http://www.georgiasouthern.edu/ijsotl/issue_v2n1.htm.

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11*(3), 203-270.

Kovac, J., & Sherwood, D. W. (1999). Writing in chemistry: An effective learning tool. *Journal of Chemical Education, 76*(10), 1399-1403.

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education, 23*(2), 157-172.

Lowman, J. (1996). Assignments that promote learning. In R. J. Menges, M. Weimer, & Associates (Eds.), *Teaching on solid ground: Using scholarship to improve practice.* San Francisco: Jossey-Bass.

Margerum, L. D., Gulsrud, M., Manlapez, R., Rebong, R., & Love, A. (2007). Application of calibrated peer review (CPR) writing assignments to enhance experiments with an environmental chemistry focus. *Journal of Chemical Education, 84*(2), 292-295.

McCarty, T., Parkes, M. V., Anderson, T. T., Mines, J., Skipper, B. L., & Greboksy. (2005). Improved patient notes from medical students during web-based teaching using faculty-calibrated peer review and self-assessment. *Acad Med*, *80*, 67-70.

Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International, 41*(3), 273-290.

Pelaez, N. J. (2002). Problem-based writing with peer review improves academic performance in physiology*. Advanced Physiology Education, 26*, 174-184.

Rivard, L. P., Stanley, B., & Straw, S. B. (2000). The effect of talk and writing on learning science: An exploratory study. *Science Education, 84*(5), 566-593.

Russell, A. (2001). *The evaluation of CPR. Prepared for HP e-Education; Business Development.* Los Angeles: UCLA.

Searby, M., & Ewers, T. (1997). An evaluation of the use of peer assessment in higher education: A case study in the school of music, Kingston University. *Assessment & Evaluation in Higher Education, 22*(4).

Sherwood, D. (1999). Writing in chemistry: An effective learning tool. *Journal of Chemical Education, 76*(10), 1399-1403.

Sluijsmans, D., Brand-Gruwel, S., Van Merrienboer, J. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education, 27*(5), 443-454.

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research,* 1, 293-319.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.

Graph Set 1
Group TRs by Instructor

Instructor "A"

Instructor "B"

Instructor "D"

Instructor "E"

Instructor "F"

Instructor "H"

Instructor "I"

Graph Set 2
Group RCIs by Instructor

Instructor "A"

Instructor “B”

Instructor “C”

Instructor “D”

Instructor "E"

Instructor "F"

Instructor "H"

Instructor "I"

Graph Set 3
Group TRs and RCIs for student in Instructor "G"'s class