

# An Analysis of Calibrated Peer Review (CPR) in a Science Lecture Classroom

By Mark E. Walvoord, Mariëlle H. Hoefnagels,  
Douglas D. Gaffin, Matthew M. Chumchal, and David A. Long

*Note:* Dr. Ann Cutler served as guest editor for this issue's Research and Teaching columns.

*Calibrated Peer Review (CPR) is an online tool being used to integrate a writing component in classrooms. In an introductory zoology lecture class, we found that CPR-assigned scores were significantly higher than instructor-assigned scores on two of three essay assignments. We also found that neither students' technical-writing skills nor their abilities to convey scientific understanding of articles through summary essays improved during the semester. The CPR system offered fairly simple setup and submission for students and decreased grading time for instructors, so we offer suggestions on CPR settings and the best usage of this program.*

Calibrated Peer Review, or CPR, is an online tool for submission and peer review of student essays. CPR was originally developed in 1998 to foster science literacy, constructivist learning, and critical thinking in introductory chemistry classrooms in California colleges and universities (Russell, Chapman, and Wegner 1998), but it has now been used by more than 525 educational institutions (The Regents of the University of California 2001) in many disciplines, class sizes, and levels of education.

CPR's widespread use is a testament to its flexibility. Instructors can create virtually any type of writing assignment in CPR, set word-count limits,

set and adjust student deadlines, create grading rubrics, control the relative weights of the components that CPR uses to calculate assignment grades, adjust the grading difficulty, view student progress, and, when necessary, override student scores. Moreover, instructors can use or modify assignments that other users have deposited in CPR's collection.

Given CPR's flexibility and underlying complexity, a complete overview of its workings is beyond the scope of this paper (see the white paper at <http://cpr.molsci.ucla.edu> and Robinson 2001 for more information). In brief, each assignment consists of four main stages: essay submission, calibration, peer review, and self-assessment. The instructor sets the time window during which students submit their work to CPR, after which the calibration stage begins. Students use an instructor-written grading rubric to evaluate three "calibration essays" (poor, average, and excellent answers to the assignment) presented in random order. As students complete the rubric for each calibration essay, CPR compares their responses to the ideal answers set up by the instructor. The number of student answers matching the ideal answers provides a measure of each student's grading proficiency, or reviewer competency index (RCI), which later factors into the student's overall assignment grade. After the calibration stage, students use the same grading rubric to answer questions about three anonymous essays written by their peers, assigning each a score

from 1 to 10. The fourth assignment stage, self-assessment, requires that students again use the grading rubric to score their own essays.

Each student's grade includes four components for which the instructor decides the relative weightings. The first component is the weighted average of three peer scores of the student's essay. In calculating this average, CPR weights the high-RCI students' evaluations more heavily than those of graders with low RCIs. The second component is the student's performance on the calibrations. The third component is a comparison of the scores each student assigned to each of three peer essays with the average scores two other reviewers gave the same essays. The fourth component is a comparison of each student's self-assessment score to the average ratings from the three graders who scored the student's essay. The latter two components derive from deviations from the average, and the instructor can set allowable deviations for full, half, and no credit.

One premise of peer review is that it helps students hone their writing skills through evaluating other students' essays and comparing them to their own (Liu et al. 2001; Sims 1989). The main rationale for peer review is to improve student writing, conceptual understanding, and critical thinking. We found no peer-reviewed publications, however, to support the hypothesis that CPR improves writing skills, and we found little published research on either student attitudes toward CPR or

how a CPR-calculated grade compares to an instructor's grade on the same assignment (but for a general discussion of this latter point, see Falchikov and Goldfinch 2000). Robinson (2001) reported that students' overall attitude toward CPR was poor, but suggested that it might improve if students received extra credit for learning the system on a first assignment. Robinson also observed that complaints declined with subsequent use of CPR. In a large (1,100-student) class, we observed that clear presentation of the procedures and positive instructor attitude may help improve student attitudes (Douglas D. Gaffin and Mark E. Walvoord, unpublished data).

In this paper, we evaluate the use of CPR in a first-year biology course at the University of Oklahoma. The research presented here had three main objectives: (1) to compare scores assigned through CPR versus instructor-assigned grades; (2) to determine if technical-writing skills and scientific understanding improved throughout the semester while using CPR; and (3) to assess student opinions on ease of use, usefulness, and fairness of CPR.

## Materials and methods

Between 54 and 59 students in a 2004 summer section of Introductory Zoology at the University of Oklahoma (OU) used CPR to submit one to four essays. The first was a "training" essay designed primarily to teach the use of CPR, and the other three were substantive assignments requiring written summaries of scientific articles. Table 1 summarizes the CPR settings we used.

After students completed each of the three substantive assignments, a random sample of 20 student essays was selected (a new random sample for each of the three essays, with some students represented more than once because of class size). Once names were removed from the es-

says, three biology instructors who primarily teach introductory zoology/biology (Douglas D. Gaffin, Mariëlle H. Hoefnagels, and Mark E. Walvoord—instructor of this course) graded them using the same rubrics as students used. Each rubric included nine yes/no questions on content and one yes/no question on spelling and grammar (Table 2). Each essay was rated on a scale of 1 to 10 by adding up the number of "yes" answers. For each student's essay, the average instructor score was compared to (1) the text quality scores assigned through peer review and (2) the overall score assigned by CPR. This second comparison incorporated not

only the average text rating of the two or three peer reviewers based on the 10-question rubric, but also the self-assessment, calibration, and review-of-others scores. We used paired-sample *t*-tests (SPSS version 12.0) to compare these scores.

To determine whether the technical quality and scientific understanding of student writing improved throughout the semester, we randomly selected 20 students (a new random selection, independent of the three sets above) who had completed all essay assignments and had not asked for any essay re-grades. We provided the three essays for each of these 20 students to a writing expert (David A. Long,

**TABLE 1**

**Calibrated Peer Review settings for essay assignments one, two, three, and four.**

Word limit	250, 400, 450, and 450 respectively
Assignment scoring	Low difficulty <sup>a</sup>
Grade breakdown	Text quality: 60% (as scored by their peers) Calibration performance: 10% Reviews of peer essays: 20% Self-assessment: 10%
Timing	Essay submission: Approximately five days Calibrations, peer essay reviews, and self-assessment: Approximately five days
Student explanation	Not required <sup>b</sup>

<sup>a</sup> Students could deviate by 3.0 or fewer points from the rating of calibration texts or the average rating of peer-reviewed essays and still receive full credit for these portions; students could deviate by 2.0 or fewer points from the average rating of their own text to receive full credit or between 2.0 and 3.0 points to receive half credit on this portion (compared to "moderate-" or "high-difficulty" scoring).

<sup>b</sup> Students were not required to explain their scoring of each rubric question for peers' essays.

director of OU's Expository Writing Program) to grade for technical writing improvement, and to a zoology doctoral candidate and instructor (Matthew M. Chumchal) to grade for scientific understanding. Neither were

given indication of the students' identities or the order in which they wrote their essays. Long was chosen because of his job position and his lack of background in biology. He scored the 60 essays in random order for

technical quality, using two 10-point scales of evaluation: one for "usage" (grammar and style, including spelling and typographical errors) and one for "coherence" (clarity and cogency of expression, including vagueness, redundancy, economy, organization, and continuity). Thus, each essay achieved a technical quality score out of 20 possible points. Chumchal similarly scored each essay using a rubric to check for correct identification of hypotheses, methods, data collected, interpretation of results, and support of hypotheses with a total possible score of 10 points. We analyzed the overall technical-writing scores, the individual score components of "usage" and "coherence," and the scientific-understanding scores with the univariate approach to the general linear model (SPSS version 14.0) to determine if either writing skills or scientific understanding changed significantly over time.

To assess student opinions about CPR, we distributed four anonymous, in-class questionnaires soon after each assignment was complete. The questionnaire distributed after the training essay differed from the other three. It asked students to rate four statements ("The process of creating an account in CPR was simple" plus questions 1 through 3 from the subsequent questionnaires; see Figure 1). Following each statement was space for written explanations, and the end of the survey provided space for miscellaneous comments. For the subsequent three questionnaires, students rated nine statements in the questionnaires using the same seven-option scale (see Figure 1 note), followed by a space for written explanations. We also included two open-ended questions about the main advantages and disadvantages of CPR.

The first questionnaire was distributed to students with a cover letter explaining the purpose of the study. Both through this letter and orally

**TABLE 2**

**Sample grading rubric for essays 2–4.**

	Rubric questions	Answer choices
1	Does the summary state that the objective of this study was to determine how eggshell pattern/pigment is inherited (and/or what the purpose of this pattern/pigment is)?	Yes No
2	Does the summary state at least two pieces of background information about eggs, pigmentation, and/or inheritance of egg color?	Yes No
3	Does the summary state the hypothesis that egg pigmentation is under genetic control through the female line? (OR on the W chromosome?)	Yes No
4	Does the summary state that the study subject was the great tit ( <i>Parus major</i> ) or the Wytham population of birds? AND does the summary further state that the sample size was 1,104 (egg) clutches, 654 female moms, or 863 identified clutches?	Yes No
5	Does the summary state the type of data collected (at least one of the following)? <i>Eggshell spot patterns OR more specifically, spot intensity, distribution of spotting, and average spot size OR measurements of parents (age, sex, weight, length, etc.)</i> AND does the summary state at least one of the following statistics used? <i>Shapiro-Wilks' tests, intraclass correlation, OR general linear models</i>	Yes No
6	Does this summary state at least one key finding of the research paper?	Yes No
7	Does the summary state that the research supported the hypothesis?	Yes No
8	Does the summary state some way that this research paper relates to what we've been covering in class?	Yes No
9	Does this summary include some implication of this study on the author, us, or science as a whole?	Yes No
10	Does the essay contain three or fewer spelling and/or grammatical mistakes?	Yes No
11	How would you rate this text?	Scale of 10 (highest) to 1 (lowest)

*Note:* Students were asked to read and summarize a journal article (in this case, Gosler, A.G., P.S. Barnett, and S.J. Reynolds. 2000. Inheritance and variation in eggshell patterning in the great tit *Parus major*. *Proceedings of the Royal Society of London*, B, 267 (1461): 2469–73). They were asked to include (1) purpose/hypothesis of this study, (2) background information, (3) materials/methods used, (4) results of the study, and (5) discussion/conclusion points. Students then used the rubric above to score calibration essays, peer essays, and their own essays.

in class, the instructor explained that neither course grades nor essay scores would be affected by participation or nonparticipation in the questionnaires. Moreover, students were assured of the anonymity of their questionnaire responses and the confidentiality of essay scores in the event of publication of this research. (Note: This research was performed under approval from the Institutional Review Board at the OU-Norman Campus [FWA #00003191]).

## Results

### Overall student performance and comparisons to instructor scores

The assignment completion rate over the four assignments ranged from 83% to 95%, and approximately 5% of students turned in text but failed to complete calibrations or peer reviews. Response rates on post-essay questionnaires varied from 35% to 66% (Table 3).

Average instructor-given scores ( $\pm$  standard deviation) were significantly lower than CPR-given scores for the second ( $7.9 \pm 1.6$  vs.  $8.9 \pm 1.1$ ;  $t_{19} = -4.25, p < .001$ ) and third essays ( $8.1 \pm 1.8$  vs.  $8.6 \pm 1.6$ ;  $t_{19} = -2.19, p = 0.04$ ), but were not significantly different for the final essay assignment ( $8.8 \pm 0.9$  vs.  $9.0 \pm 0.8$ ;  $t_{19} = -1.62, p = 0.122$ ). However, the instructor-assigned scores were very similar to the peer-review scores (i.e., scores excluding performance on calibration essays, peer review, and self-assessment), with differences between scores for the second ( $7.9 \pm 1.6$  vs.  $8.4 \pm 1.4$ ;  $t_{19} = -1.82, p = 0.085$ ), third ( $8.1 \pm 1.8$  vs.  $8.1 \pm 2.1$ ;  $t_{19} = -0.002, p = 0.998$ ), and fourth essays ( $8.8 \pm 0.9$  vs.  $8.5 \pm 1.1$ ;  $t_{19} = 1.54, p = 0.14$ ) all being insignificant (Figure 2). Most assignment grades therefore reflected a generally high level of performance on calibration essays, peer review, and self-assessment. In fact, only 2 essays of the 60 analyzed received negative consequences in scoring due to these factors.

### Technical-writing skills and scientific understanding of students through the semester

Average student technical-writing scores  $\pm$  standard deviation (usage plus coherence, see Materials and Methods) were  $15.5 \pm 3.3$ ,  $14.7 \pm 3.0$ , and  $14.9 \pm 3.4$  out of 20 for the three essays. There was no significant change in overall technical-writing skills ( $F_{2,38} = 1.034, p = 0.366$ ), usage scores ( $F_{2,38} = 1.179, p = 0.319$ ), or coherence scores ( $F_{2,38} = 0.266, p = 0.768$ ) throughout the semester.

Student's mean scientific-understanding scores (out of 10 points) on the three essays  $\pm$  standard deviation were  $6.9 \pm 1.2$ ,  $7.1$

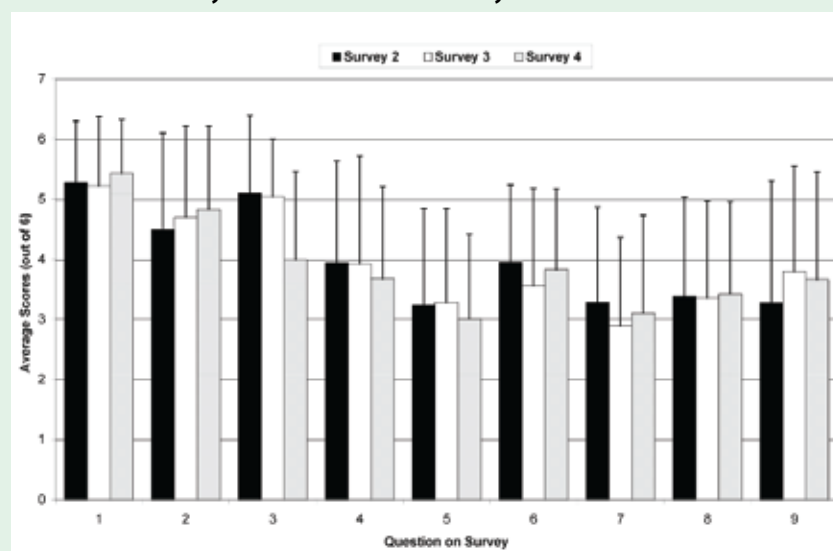
$\pm 1.3$ , and  $6.6 \pm 1.3$  respectively. There was no significant change in overall scientific understanding during the semester ( $F_{2,38} = 0.939, p = 0.400$ ).

### Student perceptions of CPR

The questionnaires showed that students agreed that submitting assignments, the calibration step, and the peer-review step were simple in CPR. They also agreed that the final results were easy to understand. On average, they "agreed more than disagreed" with their final grade, that writing improved their critical reading, and that CPR is preferred over turning in a regular paper. Students "disagreed

FIGURE 1

Results from surveys administered after essays 2–4.



Note: Histogram shows mean scores + standard deviation. The statements rated by students on surveys after completion of essays 2–4 were as follows: (1) The process of submitting an assignment in CPR was simple, (2) The processes of calibration and peer review were simple, (3) The assignment results (CPR's explanation of how your grade was determined) were easy to understand, (4) I agree with the grade I received on my CPR assignment, (5) Writing for this assignment helped me better understand the related course material, (6) Writing for this assignment helped me improve my critical reading skills (Note: One definition of *critical reading* is "the process of making judgments in reading: evaluating relevancy and adequacy of what is read"), (7) The calibration and peer-review portions of this assignment helped me better understand the related course material, (8) The calibration and peer-review portions of this assignment helped me improve my critical reading skills (see question 6 for definition of critical reading), and (9) Considering this method as a whole, I prefer CPR over turning in a "regular" paper to my instructor.



more than agreed” that the CPR process helped them understand the related course material and that calibration and peer review helped

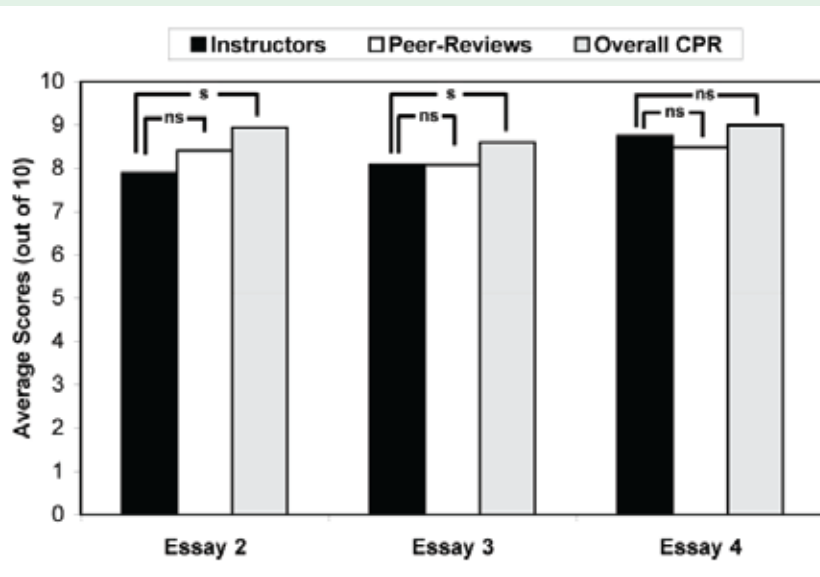
them read critically (Figure 1). In written comments collected on the questionnaire, approximately 10% of respondents complained about the

time necessary to complete calibrations and reviews, and approximately 10% expressed frustration that their grades depended on their peers’ opinions. Less commonly, students speculated that CPR might “dumb down” writing, because students might write in a simplistic form tailored to the grading rubric. Other students observed that some peer graders were more thorough than others, and they worried that their grade might be negatively affected if peers did a good job on the calibrations (got a high RCI), then rushed through the peer-review stage and incorrectly rated their essay.

Positive comments included that the system was convenient (built-in word count, online submission, flexible timing for calibrations/reviews); that calibrations helped students compare their own essays to good ones (i.e., the instructor’s “high-quality” answer); that reviews let students see diverse writing styles and ways of conveying the same ideas; and that CPR probably assigned higher grades than the instructor would have.

**FIGURE 2**

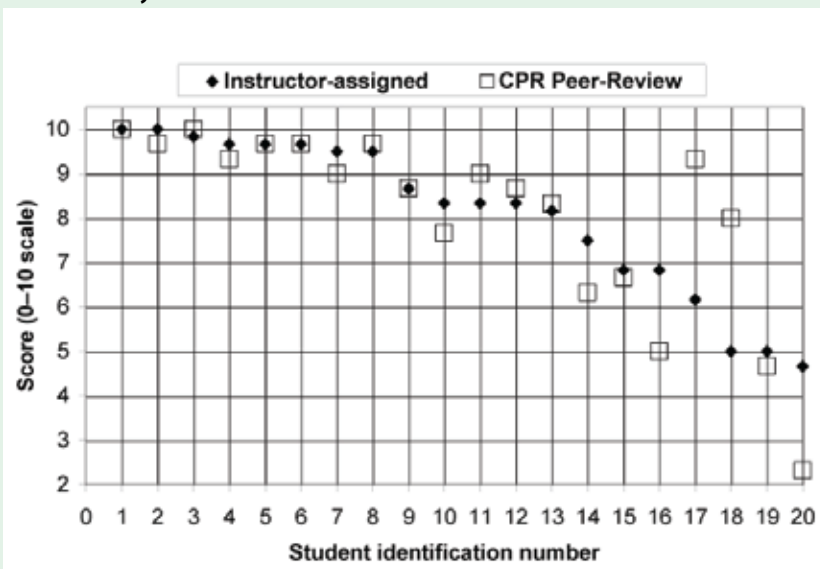
Essay scores using three methods.



Note: Pairs of scores labeled “s” showed significant difference, while pairs labeled “ns” showed no significant difference.

**FIGURE 3**

Average instructor score versus average CPR peer-review score for 20 students on the third essay.



Note: Students rated their opinion from “1, strongly disagree” to “6, strongly agree” or “No opinion.” “No opinion” scores were omitted from analyses.

## Discussion

Our findings show that in its current form and with the settings listed (Table 1), CPR-given scores could be significantly different than those conferred by instructors. Our findings further show that CPR did not appear to increase students’ technical-writing abilities over the span of the semester, nor did it improve their ability to relay increased scientific understanding of published articles through written summaries. Questionnaires show that students generally agreed with statements about the usefulness of CPR, but negative student comments on the open-ended survey questions outnumbered positive ones.

High CPR scores could be due to a “training lag” as students learned how to grade correctly using rubrics, empathy of students

toward other students' grades, or just a general lack of attention to detail by students. Different CPR sub-score weightings could have reduced discrepancies between CPR-assigned student scores and instructor-assigned scores, since overall CPR-assigned scores depend heavily on the instructor-specified grade breakdown (see Table 1). Had we weighted the scores at 100, 0, 0, and 0% respectively for text quality, calibration performance, reviews of peer essays, and self-assessment review (i.e., used only the peer-reviewed score of text quality and no other factors for the final CPR-assigned score), the scores assigned by CPR would not have been significantly different than instructor-assigned scores. However, the latter three parameters may need to be weighted to some degree to encourage students to perform these tasks. The system default is 20, 30, 30, and 20% for the weightings, but this breakdown would have elevated our student scores even higher above instructor-assigned determinations. On the other hand, weightings of 80, 5, 10, and 5% might produce CPR-assigned grades closer to instructor-assigned grades, while still providing some incentive for students to perform well on the supplemental duties of the system (as long as students perceived 5, 10, and 5% to be adequate incentive for the supplemental duties).

While average CPR peer-review scores (versus overall CPR scores) were not significantly different from instructors' scores, individual students might have concerns about fairness when they compare their results to their peers, leading to complaints or calls for re-grades. For instance, for the third essay assignment, the mean instructor-determined score was equal to the CPR peer-review score of 8.1 for the 20 randomly selected students (Figure 2). However, two students (students #18 and #19, Figure 3) had essays of similar quality according to the instructor-assigned scores using the rubric, but they received disparate reviews from peers. Student #19 received a 33.3% lower CPR peer-review score, which translated to a 30% overall CPR score reduction. This discrepancy appears to be due to variation in competence among reviewers, but was tempered by the affected student's performance on the three supplemental CPR components (since those had a certain percentage weight of their overall CPR grade). Another method to combat this problem would be to allow more than three reviewers per essay. However, currently CPR doesn't allow any changes to this part of the system, and having to review more than three other essays would only increase the amount of time students spend on each assignment (already a source of negative comment).

CPR did not appear to improve student technical writing or scientific understanding for the assignments in this study. The best writers in the class may have read only essays that were more poorly written than their own, and they would not have been helped, except perhaps through the best calibration essay. However, we would have expected that a lack of improvement by "good" writers should have been offset by improvement by poor writers. The poorest writers had the most room for improvement through their calibration and review of essays, many of which would be better than their own. Yet even the worst writers might not improve, because the rubric might encourage students to write simplified essays that incorporate just the necessary information. A more beneficial way to use the CPR system would be to require comments on the peer reviews, then to require a second round of essay submission with the same writing prompt, where students could incorporate what they learned on the first round to write a technically better and more scientifically accurate essay.

Our instructions directed students to count the "yes" scores on the rubric to determine the answer to the "How would you rate this text?" question in the CPR rubrics. Some students performed this step improperly; for example, a student answering "yes" to 8 of 10 questions

**TABLE 3**

**Number and percentage of student participants for each essay.**

Essay #	Did not start assignment	Started but did not finish assignment	Finished assignment	Responded to survey after assignment	Total students enrolled at time of assignment
1	5 (8%)	2 (3%)	52 (88%)	39 (66%) <sup>a</sup>	59
2	5 (8%)	5 (8%)	49 (83%)	22 (37%)	59
3	0 (0%)	3 (5%)	53 (95%)	28 (50%)	56
4	3 (6%)	4 (7%)	47 (87%)	19 (35%)	54

<sup>a</sup> Misprinted scale on questionnaire; so we used only those 19 student surveys that used the corrected scale given in class

on the rubric might score the essay as a five. In some of these cases we believe that students intentionally manipulated the assigned score because they weren't satisfied that the rubric-derived score was reflective of the quality of the essay. This could indicate (along with written comments on our surveys) that students weren't fully satisfied with a simple yes/no rubric, but rather preferred to have some subjective component of rating to satisfy their intuitive sense of "good writing." One result of including a more subjective measure might be an increase in "coherence" scores over a semester that includes multiple CPR assignments. Another strategy to increase coherence might be to require students to explain their reasoning in assigning each peer grade.

Although usage of the CPR system didn't appear to increase writing quality, other CPR users have found positive results, both anecdotally and through published research. One instructor's students have returned to him a semester or more after using CPR in his classroom to let him know they think CPR really helped them write better in subsequent classes (Todd Nickle, Mount Royal College, Alberta, Canada, pers. comm.). In addition, using CPR with slightly different settings may increase writing skill and critical thinking (Heise, Palmer-Julson, and Su 2002). Contrary to our findings that scientific understanding did not improve, Pelaez (2001) found that human-physiology majors performed significantly better on the multiple-choice and essay portions of their midterm exams over topics taught using CPR compared to topics taught in a traditional method.

Student attitude toward CPR remains a problem, but this may be true of any writing assignment in a course. We tried to improve students' attitudes at the beginning of the course by explaining CPR

procedures and presenting preliminary data from a previous course showing that student scores were, on average, higher using peer review than if the instructor had graded the essay assignments. The presentation included information on the instructor's ability to see all comments and reviews, the instructor's ability to override any grades if necessary, the effects of poor evaluation or calibration on the individuals' scores, and the effects of a low RCI on others' text-rating scores.

Although students "slightly agreed" with the grade they received on each assignment, several written comments revealed that students did not trust their peers to determine their grades. Overall, students "slightly agreed" that using CPR was better than the traditional method of turning in a paper to the instructor. This positive result could be due to the use of technology in the classroom, enthusiasm of peers or the professor, or preconceived notions about the system.

To help improve student attitudes and inequalities among students, a system of appeal or re-grade should be implemented. Also, instructors should use caution when using CPR to assign grades until they have enough experience with the system to understand the settings that work best for their class, to write rubric questions that are concise and cover all important criteria for the writing assignment, and to craft writing prompts that support course goals. The best uses of CPR may be for assigning questions requiring more objective answers (especially in introductory classes); in coordination with group work in some hybrid setup; and/or for refining already "good" writing skills in upper-division courses (where more subjective scaling could be used).

Future studies on the CPR system should continue to measure increased student understanding over a semes-

ter, and should seek to incorporate data on longer term (multi-semester) changes in writing skills and scientific understanding compared to non-CPR users. Further, a study on the efficacy of the system using more subjective rubric questions and their effect on increased student writing would be helpful. To understand the effects of having only three reviewers per essay, a study could be conducted with more essay reviews per student (external to the CPR system), comparing the groups for attitude and overall score (versus instructor-given scores).

CPR's strong points appear to be (1) ease of use for students; (2) allowing a computer-based versus paper-based assignment that incorporates technology in the classroom, ensures confidentiality, and eases grade compilation (Sung et al. 2005); (3) the opportunity for students to see their peers' work without names attached; (4) fairly accurate diagnosis if a good rubric is provided; and (5) reduced time grading essays for the instructor.

#### Acknowledgments

*The authors would like to thank the students who took part in this study, and the Institutional Review Board of our institution for permission to perform this study. In addition, we are grateful to the anonymous reviewers of this manuscript and to CPR for providing and supporting a free educational resource that we could evaluate in our classroom.*

#### References

- Falchikov, N., and J. Goldfinch. 2000. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research* 70 (3): 287–322.
- Heise, E.A., A. Palmer-Julson, and T.M. Su. 2002. Calibrated peer review writing assignments for introductory geology courses. *Abstracts with programs (Geological Society*

- of America*) 34 (6): A-345.
- Liu, E.Z., S.S.J. Lin, C. Chiu, and S. Yuan. 2001. Web-based peer review: The learner as both adapter and reviewer. *IEEE Transactions on Education* 44 (3): 246–51.
- Pelaez, N.J. 2001. Calibrated peer review in general education undergraduate human physiology. In *Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science (Costa Mesa, California, 18–21 January, 2001)*. eds. P.A. Rubba, J.A. Rye, W.J. DiBiase, and B.A. Crawford, 1518–30.
- The Regents of the University of California. 2001. Who uses CPR? [http://cpr.molsci.ucla.edu/who\\_uses\\_cpr.asp](http://cpr.molsci.ucla.edu/who_uses_cpr.asp).
- Robinson, R. 2001. An application to increase student reading and writing skills. *The American Biology Teacher* 63 (7): 474–80.
- Russell, A.A., O.L. Chapman, and P.A. Wegner. 1998. Molecular science: Network-deliverable curricula. *Journal of Chemical Education* 75 (5): 578–79.
- Sims, G.K. 1989. Student peer review in the classroom: A teaching and grading tool. *Journal of Agronomic Education* 18 (2): 105–08.
- Sung, Y-T., K-E. Chang, S-K. Chiou, and H-T. Hou. 2005. The design and application of a web-based self- and peer-assessment system. *Computers and Education* 45 (2): 187–202.

**Mark E. Walvoord** (walvoord@ou.edu) is the assistant director of the Assessment and Learning Center, **Mariëlle H. Hoefnagels** is assistant professor of zoology and botany/microbiology, **Douglas D. Gaffin** is dean of University College and associate professor of zoology, and **David A. Long** is director of the Expository Writing Program, all at the University of Oklahoma in Norman. **Matthew M. Chumchal** is assistant professor of biology at Texas Christian University in Fort Worth, Texas.



# KNOPF

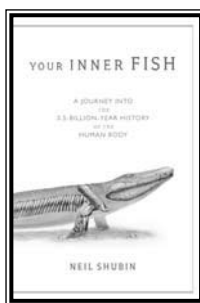
ALFRED A. KNOPF ♦ VINTAGE ♦ ANCHOR BOOKS  
PANTHEON ♦ SCHOCKEN ♦ EVERYMAN'S LIBRARY



*George Johnson*  
**THE TEN MOST BEAUTIFUL EXPERIMENTS**

From the acclaimed *New York Times* science writer George Johnson, an irresistible book on the ten most fascinating experiments in the history of science. We see Galileo singing to mark time as he measures the pull of gravity, and Newton carefully inserting a needle behind his eye to learn how light causes vibrations in the retina. William Harvey ties a tourniquet around his arm and watches his arteries throb above and his veins bulge below, proving that blood circulates. In bringing us these and seven other dazzling stories, Johnson restores some of the romance to science, reminding us of the existential excitement of a single soul staring down the unknown.

KNOPF | CLOTH | 208 PAGES | \$22.00

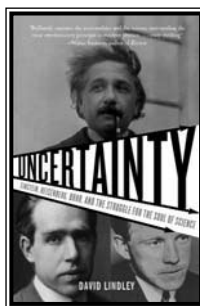


*Neil Shubin*  
**YOUR INNER FISH**

A Journey into the 3.5-Billion-Year History of the Human Body

"I was hooked from the first chapter of *Your Inner Fish*. Creationists will want this book banned because it presents irrefutable evidence for a transitional creature that set the stage for the journey from sea to land. This engaging book combines the excitement of discovery with the rigors of great scholarship to provide a convincing case of evolution from fish to man." —Don Johanson, director, Institute of Human Origins, Arizona State University

PANTHEON | CLOTH | 240 PAGES | \$24.00



*David Lindley*  
**UNCERTAINTY**

Einstein, Heisenberg, Bohr, and the Struggle for the Soul of Science

"Brilliantly captures the personalities and the science surrounding the most revolutionary principle in modern physics. This book is . . . truly thrilling."

—Walter Isaacson, author of *Einstein: His Life and Universe*

"Far and away the best popular account of the development of quantum mechanics I have encountered."

—Michael D. Gordin, *American Scientist*

ANCHOR | PAPER | 272 PAGES | \$14.95

KNOPF ACADEMIC ♦ RANDOM HOUSE, INC.  
1745 BROADWAY, 20-2 ♦ NEW YORK, NY 10019  
WWW.RANDOMHOUSE.COM/ACADEMIC  
ACMART@RANDOMHOUSE.COM