

## Article

# Writing Assignments with a Metacognitive Component Enhance Learning in a Large Introductory Biology Course

Michelle Mynlieff, Anita L. Manogaran, Martin St. Maurice,  
and Thomas J. Eddinger

Department of Biological Sciences, Marquette University, Milwaukee, WI 53233

Submitted May 22, 2013; Revised January 24, 2014; Accepted February 16, 2014  
Monitoring Editor: Mary Pat Wenderoth

Writing assignments, including note taking and written recall, should enhance retention of knowledge, whereas analytical writing tasks with metacognitive aspects should enhance higher-order thinking. In this study, we assessed how certain writing-intensive “interventions,” such as written exam corrections and peer-reviewed writing assignments using Calibrated Peer Review and including a metacognitive component, improve student learning. We designed and tested the possible benefits of these approaches using control and experimental variables across and between our three-section introductory biology course. Based on assessment, students who corrected exam questions showed significant improvement on postexam assessment compared with their nonparticipating peers. Differences were also observed between students participating in written and discussion-based exercises. Students with low ACT scores benefited equally from written and discussion-based exam corrections, whereas students with midrange to high ACT scores benefited more from written than discussion-based exam corrections. Students scored higher on topics learned via peer-reviewed writing assignments relative to learning in an active classroom discussion or traditional lecture. However, students with low ACT scores (17–23) did not show the same benefit from peer-reviewed written essays as the other students. These changes offer significant student learning benefits with minimal additional effort by the instructors.

## INTRODUCTION

The ability of an instructor to facilitate student learning, and specifically to enhance his or her students’ ability for higher-order application, analysis, and problem solving, is essential for the rapidly advancing fields of science, technology, engineering, and math (STEM). One of the critical components to facilitate learning is the development of metacognitive skills (Schraw, 2002). Metacognition is the ability of a student to actively regulate his or her learning. This requires students to recognize their own mistakes and monitor their level of

understanding (Gourgey, 2002). Throughout their four years of college, successful students will gain skills in metacognition, and the ability to monitor their own cognitive ability will help them to succeed in future endeavors (Schraw, 2002). In the present study, we sought to determine whether two individual metacognitive tasks enhance performance on subsequent assessments that require critical thinking. In addition, students entering as freshmen are likely to have different metacognitive abilities that would be reflected by their previous academic performance, and one might expect that tasks designed to increase metacognition may differentially affect populations of students with varying academic abilities. The students in the present study were disaggregated based on ACT scores to identify differences in their response to the tasks that included metacognition.

Many studies have investigated the ability of writing to enhance learning (for a review, see Bangert-Drowns *et al.*, 2004). The reported effectiveness of using writing-to-learn strategies within the classroom varies greatly depending on a number of factors. Reynolds and colleagues (2012) performed an exhaustive literature review of studies published after 1994 on writing strategies specifically used in STEM

DOI: 10.1187/cbe.13-05-0097

Address correspondence to: Thomas J. Eddinger (thomas.eddinger@marquette.edu).

© 2014 M. Mynlieff *et al.* CBE—Life Sciences Education © 2014 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

disciplines. Examining 324 journal articles, books, book sections, conference proceedings, and reports, they found that most were descriptive case studies. Many articles describe interesting assignments that can be used in the classroom but do not rigorously test the effectiveness of these assignments in enhancing learning. In addition, studies often rely on student perception as an assessment rather than a rigorous testing of student critical-thinking abilities (Clase *et al.*, 2010; Brownell *et al.*, 2013). Reynolds and colleagues highlight the need for studies that test the effectiveness of writing-to-learn strategies.

The type of writing assignment is crucial in enhancing different aspects of learning (Durst and Newell, 1989; Bangert-Drowns *et al.*, 2004). The act of taking notes and writing brief essays on content are likely to lead to increased retention due to increased exposure to the material (time on task) but may not enhance critical thinking (Weinstein and Mayer, 1983). While these tasks are particularly suited to enhancing declarative knowledge, it is imperative that today's students gain skills in higher-order critical thinking and the ability to apply concepts to new situations. With the advent of the Internet, information is immediately available through computers and other smart devices. However, successful workers in the 21st century need to be capable of processing a wealth of information in order to analyze problems and new situations. To help students gain these skills, writing tasks should have both an analytical component and a metacognitive component, requiring both analysis and reflection (Paris and Paris, 2001; Zumbrunn *et al.*, 2011).

Writing assignments are a natural fit within smaller upper-division classes. With a small class, the professor has the opportunity to provide quality feedback on written assignments to students, and assignments often include analysis of material from different sources, multiple editing steps, and reflection. Furthermore, these classes are primarily populated by students majoring in the subject who may be more motivated to learn the material in depth than students in lower-division courses (Ainley *et al.*, 2002). Conversely, providing analytical writing opportunities is difficult in large introductory classes, particularly if there is a limit on teaching assistants. In the present study, we designed two different writing tasks that incorporate metacognitive features and yet are reasonable to execute in a large-lecture setting. Notably, we took advantage of the three sections of General Biology 1001 that are offered at Marquette University and used a scientific design to analyze the effects of these writing tasks on enhancing student learning. Specifically, we coordinated assignments across the three sections to design, test, and quantitatively evaluate objective measures of the effectiveness of these writing interventions, both across the entire student population and in disaggregated populations based on ACT scores.

The first intervention analyzed in the present study was the use of exam corrections as a learning tool. Exams are almost universally included in all courses to assess student progress and accomplishment. While exams often serve as assessment tools, their ability to serve as a learning tool may be underappreciated by instructors and students alike (Harper *et al.*, 2004; Yerushalmi *et al.*, 2007; Roediger and Karpicke, 2006; Roediger and Butler, 2011). In large introductory science courses, exams are primarily multiple choice. Metacognition and reflection on incorrectly answered questions can be the difference

between knowledge (knowing, familiarity, awareness) and understanding (comprehension, discernment, perceived significance) of course material and can serve to correct student misconceptions. Henderson and Harper (2009) documented their use of exam corrections in encouraging students to reflect on their mistakes, but their study did not directly control for other features of their pedagogy, nor were the students disaggregated based on ability.

The second intervention involved writing essays using the Calibrated Peer Review (CPR) system developed at University of California–Los Angeles (UCLA; Russell *et al.*, 1998; <http://cpr.molsci.ucla.edu>). At first glance, the program appears to be designed to allow faculty with large classes and limited help from teaching assistants to assign and grade essays, because the “grading” is performed by the students through peer review. However, CPR offers much more than a simple mechanism for grading an assignment in high-enrollment lecture courses; the review portion of the assignment adds a strong metacognitive element. The up-front work of constructing a well-designed assignment with a clear grading rubric is crucial to the success of the system. The instructor can dictate the extent of critical thought and analysis required for a given assignment. The system has been used to have students write straightforward essays on class content, summarize scientific articles (Walvoord *et al.*, 2007), and work through problems or analyze scientific issues after reading multiple sources (Pelaez, 2002; Libarkin and Ording, 2012). After the initial submission, the student must grade three calibrated essays and then three anonymous essays from their classmates using the provided rubrics. The final task is to review their own essays based on the rubric and in light of the six other essays they have reviewed. It can be envisioned that as much “learning,” if not more, takes place during the review phase as in the initial writing phase. Once the assignment is written using the computer-based system, it can be used in any size class, so it lends itself well to a large-lecture format.

The goal of this study was to experimentally test whether writing interventions that have metacognitive components offer academic benefits to students. Taking advantage of our current three-section, three-instructor introductory biology course, we designed an experimental study to test the possible benefits of 1) postexam analysis and 2) peer-reviewed writing assignments using control and experimental variables across and between sections. Our data show that students who participated in structured independent written corrections or postexam analytical group discussion performed better on subsequent assessments than those students who did not. In addition, students who participated in writing assignments performed better on a subsequent short-term (~2 wk postactivity) or longer-term (at least 4 wk postactivity) assessment than those who learned the material in a small-group discussion or a traditional lecture (passive learning). Students in the low ACT group benefited more from the postexam analysis than students in the midrange to high ACT groups, whereas the students in the low ACT group benefited the least from the written essay assignment. These results indicate that the use of well-designed writing interventions that include a metacognitive component lead to enhancements in critical-thinking abilities in a large introductory course setting.

**Table 1.** Demographics of students in General Biology 1001

College	Number of students
Arts and Sciences	256
Health Sciences	253
Engineering	119
Business Administration	16
Communications	13
Education	12
Nursing	7
Professional Studies	1
Year in school	
Freshman	593
Sophomore	57
Junior	21
Senior	6
Gender	
Female	394
Male	283
Total students in study	677

## METHODS

### *Course Background and Student Demographics*

General Biology 1001 is the first semester of a two-semester, introductory-course series. The course is required for majors in biological sciences, physiological sciences, biochemistry and molecular biology, biomedical engineering, exercise science, and biomedical sciences, and serves as a science core course for a small percentage of non-science majors. The typical enrollment in the course is between 600 and 700 students, who are distributed across three lecture sections taught by three different full-time faculty members. Content, pedagogy, assignments, and grading policies are kept very similar across all three sections through close communication and coordination among the instructors. The major and the demographics of students enrolled in the Fall of 2011 who participated in this study are indicated in Table 1. The table includes students who withdrew from the class after late registration but before the deadline for withdrawal (11 wk).

### *Course Design in Fall 2011*

All three sections of General Biology 1001 used the same syllabus, covered the same topics, and had identical due dates for major assignments. Students met with a professor for a total of 150 min/wk (three 50-min lectures) in a large amphitheater, lecture hall setting. Lecture format varied from day to day and included traditional lecturing and small-group and clicker activities. For each course section, four 50-min discussion periods were scheduled each week that were attended by 50–60 students. Each lecture section was assigned a graduate teaching assistant who led all four discussion periods for a particular lecture section. The discussion periods were utilized for group activities, discussions, reviews, and some exam corrections. Each graduate teaching assistant was mentored by the professor responsible for his or her lecture section and given careful guidance on the preparation and execution of the discussion period activities.

Grades were based on the parameters outlined below.

**Multiple-Choice Exams (70%).** Five multiple-choice exams were administered throughout the semester, each covering 2–3 wk of material. For each exam, there were 50 questions that consisted of a combination of both lower-order and higher-order questions (Bloom, 1984). The lowest exam grade was dropped in the final grade calculation for each student. In addition, there was a cumulative final (also 50 questions) in which 50% of the questions were from midsemester exams that were unique to each section. The other 50% of the questions were new questions that were shared between all three sections, and a portion of these questions were used for assessment in this study as detailed below in *Written Essay Assignments using CPR*.

**Written Essays (9%).** Two essays were assigned and graded through CPR.

**In-Class Quizzes (6%).** Unannounced in-class quizzes were administered approximately once every other week with the use of the i>clicker classroom-response system. Three of these were also used as part of the assessment for the present study.

**Homework (5%).** Homework was assigned approximately once a week. The homework assignments primarily consisted of online tutorials with questions on the book website or written assignments posted on the Desire2Learn website for the course. There were three online quizzes administered through Desire2Learn as part of the assessment for the present study.

An additional 10% of the semester grade came from reading quizzes, participation in activities and clicker questions, and attendance.

### *Exam Corrections*

Following each of the first three exams, individual sections were assigned one of three activities in rotation: 1) written exam corrections, 2) small-group discussion of exam questions, or 3) comparison group not participating in formal exam corrections. The order in which each section participated in each activity is shown in Table 2. For the written corrections, students were given the answer key and their exams before the corrections were due, so they knew exactly which questions they answered incorrectly and the correct answers. For each incorrect answer, the students were required to explain why their answer was incorrect and why the correct answer was correct. In addition, they were required to list where they found the information (text page number, date of lecture notes, etc.). The amount of work required of each student was proportional to their original performance on the exam; that is, a student who missed two questions only had to correct two questions, whereas a student missing 15 questions had to correct all 15 questions in order to receive credit. With written corrections, students could earn back up to 20% (0–20% based on performance) of the points they missed on the exam, providing motivation for participating and making a conscientious effort in this activity.

Students who were given the small-group discussion option spent one discussion period reviewing the 10 exam questions that received the fewest correct answers on the exam given in that class. Students were allowed to discuss the

**Table 2.** Exam-correction interventions used in the three sections of the introductory biology course

Exam number <sup>a</sup>	Section A	Section B	Section C
1 (week 4)	Written corrections	None	Discussion activity and clicker quiz
2 (week 7)	None	Discussion activity and clicker quiz	Written corrections
3 (week 10)	Discussion activity and clicker quiz	Written corrections	None

<sup>a</sup>The exam is indicated along with the week the exam was administered.

questions but did not have an answer key or input from the teaching assistant during the discussion period. At the end of the class period, students were quizzed on these 10 questions using i>clickers. They received up to 20% of their missed points back on the exam depending on their performance on the clicker questions (20% for all correct, 10% for half correct, etc.), which again provided an incentive for active participation. The third section did not do any formal exam corrections. Approximately 2 wk following each exam, an assessment quiz consisting of five questions relating to the material on the most often incorrectly answered questions from the previous exam was administered in class to all sections via i>clickers (see Supplemental Material A for an example). These quizzes were unannounced, and students were given five points for participating and one point for each correct answer for a total of 10 points, which provided a reward for participating and an incentive to apply themselves. The assessment questions were designed to cover the more difficult material on the exam and, thus, the concepts overlapped with the 10 questions covered in discussion. In the final analysis, data were included only for students who participated in *both* the written corrections and the small-group discussions.

### Written Essay Assignments Using CPR

CPR is an online peer-review writing program developed at UCLA. As an introduction to the mechanics of the CPR program, and to familiarize students with their learning styles and how to “succeed” in college, all three sections of the course first read an article by Dr. Robert Leamson titled “Learning (Your First Job)” ([www.udel.edu/CIS/106/](http://www.udel.edu/CIS/106/)

iyadin/07F/misc/firstJob.pdf) and participated in an online learning-style quiz developed at Diablo Valley College. Students wrote 200–400 word essays that summarized the article and discussed their learning style. After submitting their essays to CPR, students graded three calibrated essays and three of their classmates’ essays based on a rubric provided by the instructor. Finally, students assessed their own work using the same instructor-provided rubric, which allowed them to analyze their own essays and reflect on their work. For the first essay, students were guided through the calibration and review process by the instructors.

The students’ second exposure to the CPR program was used to determine whether reading and writing on a topic is a more effective means of student learning relative to learning the material in an active group discussion or attending a traditional lecture presentation on the material. The three topics were chosen to reflect concepts that are components of the first semester of introductory biology at Marquette University (Table 3). The concepts were broad and were covered in more than a single lecture of the course. The essay on energy and macromolecules required synthesis of concepts presented as part of macromolecules (week 2), chemical energy cycles (week 6), and nutrition/digestion (week 9) in the lecture and textbook. The essay on CO<sub>2</sub> and ocean acidification required synthesis of concepts presented as part of the chemistry of water and pH (week 2) and biogeochemical cycles (week 8) in lecture and the textbook. The essay on scientific design required synthesis of concepts presented in lecture and the textbook starting during the first week of class and emphasized throughout the semester.

For each topic, one section was assigned a writing activity through CPR that included peer review and self-assessment.

**Table 3.** Essay assignments

Topic	Writing assignment	Discussion activity
Energy and macromolecules	Essay modified from “A Can of Bull?” (Heidemann and Urquhart, 2005) analyzing the marketing claims of Impulse and Red Bull Sugarfree energy drinks (1000–1400 words).	PowerPoint presentation including clicker questions and group activity analyzing the marketing claims of Impulse and Red Bull Sugarfree energy drinks.
CO <sub>2</sub> and ocean acidification	Following the reading of an article on ocean acidification (Hoegh-Guldberg <i>et al.</i> , 2007) and general reading on shellfish, students were required to answer questions (300–400 words) and write an essay (500–700 words) addressing the effects of ocean acidification on shellfish populations off the coast of Maine.	PowerPoint presentation including clicker questions on ocean acidification. Demonstration of the effects of acetic acid on shells.
Scientific design	Students were required to read a published article on pheromones (Cutler <i>et al.</i> , 1998) and write an essay (700–1100 words) on the merit of the scientific design of the study based on guiding questions.	Reading of “Love Potion #10” (Holt, 2002) and discussing the questions in small groups.

**Table 4.** Experimental design for testing the impact of writing assignments (CPR) and discussion activities on student learning

Topic	Section A	Section B	Section C
Energy and macromolecules	Writing assignment and lecture (week 10)	Discussion activity and lecture (week 9)	Lecture only
CO <sub>2</sub> and ocean acidification	Discussion activity and lecture (week 8)	Lecture only	Writing assignment and lecture (week 10)
Scientific design	Lecture only	Writing assignment and lecture (week 12)	Discussion activity and lecture (week 10)

A second section participated in an active-learning exercise that was designed by the same faculty member who wrote the CPR assignment. For this activity, students participated in small-group analysis and answered clicker questions during the discussion period. This discussion period was led by a teaching assistant. Table 4 describes how individual topics were covered in the three sections throughout the semester, including dates for the discussion activity or essay completion, with “lecture only” being the control group. The lectures corresponding to these topics were spread throughout the semester, because all three topics included components covered multiple times and in different contexts throughout the semester. For example, the topic of energy and macromolecules included information from several different units, including macromolecules, energy metabolism, and digestion. A five-question online assessment was administered approximately 2 wk (short term, 13–17 d; 10-point value) after completion of the activity (see Supplemental Material B for an example); this assessment included higher-order and analytical questions and did not require memorization of the specific assignment details. For the online assessment, multiple versions of each question were utilized; questions and answers were randomized, and a time limit was enforced for each quiz in order to minimize sharing of answers among students. For each topic, two questions on the cumulative final exam (see Supplemental Material C) were used to assess long-term analytical and higher-order thinking. These were assigned a combined value of 10 points to be consistent with the short-term assessment.

Overall, the students performed best on the assessment of scientific design. Therefore, before analysis, scores on the energy and macromolecules and CO<sub>2</sub> and ocean acidification assessments were adjusted to normalize scores to scientific design. The scores were normalized as follows:

Normalized score =

$$\text{Score} \times \left( \frac{\text{Average score on scientific design assessment}}{\text{Average score on assessment for this topic}} \right)$$

### Assigning Bloom's Categories to Assessment Questions

In designing the assessments, our goal was to ask primarily higher-order analysis and application questions and some lower-order comprehension and knowledge questions. An independent biology instructor not involved in this study was asked to rate the difficulty level of the assessment questions. Assessment questions were categorized based on the Blooming Biology Tool (BBT) established by Crowe *et al.* (2008).

Approximately two-thirds of the questions were scored as application and analysis questions, and one-third were considered knowledge or comprehension questions. Two additional evaluators from outside Marquette University also assessed the questions using the BBT. They independently scored more than 60% of the questions as application and analysis questions.

### Statistical Analysis

Where indicated, data were analyzed with one-way or two-way analysis of variance (ANOVA), followed by the Holm-Sidak procedure for multiple comparisons. This allowed us to determine interactions between multiple factors and to test the effect of a specific factor on the assessment results. The Holm-Sidak comparison procedure was chosen because it is more powerful than a Tukey or Bonferonni procedure. For the exam corrections, the change in assessment scores when compared with no corrections was tested for significance using a one- or two-sample *t* test. All statistical analyses were performed with SigmaStat or GraphPad software.

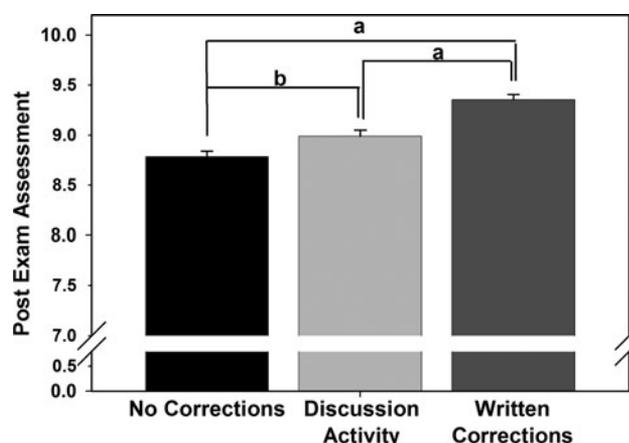
### Institutional Review Approval

All assessments and procedures involving the students were reviewed and approved by the Marquette University Institutional Review Board (HR2227). While all three sections of the course were treated differently at a single point in time and for a given topical course component, all students were afforded the same opportunity to participate in various activities and assessments throughout the semester.

## RESULTS

### Students Participating in Exam-Correcting Activities Show Increased Performance on Postexam Assessments

For testing the effectiveness of formal exam corrections, sections were assigned written corrections, small-group discussions, or no corrections (control group) in rotation for the first three exams, as shown in Table 2. For the written exam corrections, students wrote a minimum of two sentences, one explaining why their chosen answer on the exam was incorrect and one explaining the correct answer; depending on the nature of the specific question, a longer written response was sometimes given. Students who participated in written corrections or small-group discussions were given the opportunity to earn back points on their exam. Approximately 2 wk after each exam, an assessment was administered during lecture that focused on the most difficult and, thus, the most



**Figure 1.** Exam corrections increase student performance. Students participating in exam interventions (Table 2) were assessed for learning. Two weeks following each exam, five higher-order questions were administered in class for assessment (see Supplemental Material). The maximum score for each assessment was 10, with two points given for each correct answer. The assessment scores for exams 1 and 2 were normalized by multiplying each score by  $\frac{\text{Average exam 3}}{\text{Average exam 1 or 2}}$  because the assessment scores went up for all groups as the semester progressed. The normalized data are expressed as the mean  $\pm$  SEM and analyzed by a two-way ANOVA with the exam number and correction type as the independent variables followed by the Holm-Sidak method of pairwise comparisons.  $n = 447$  students; (a)  $p < 0.001$ ; (b)  $p = 0.012$ .

often incorrectly answered concepts on each exam. All students were given the option to participate in corrections and more than 80% participated in at least one correction activity. Nonparticipating students were randomly spread throughout the classes and did not correlate with exam performance, suggesting that nonparticipating students were not all of the same academic ability. To be included in the data analysis for this study, students had to participate in *all* exam correction activities offered for the first three exams. Therefore, only 67% (447 students) were included in this portion of the study.

Students' scores on the postexam assessments increased as the semester progressed, so the assessment scores for exams 1 and 2 were normalized by multiplying each score by  $\frac{\text{Average exam 3}}{\text{Average exam 1 or 2}}$ . Students assigned written corrections performed better on postexam assessments than students assigned small-group discussions (Figure 1,  $p < 0.001$ ). Of the three groups, the control groups (those who were not given a formal opportunity for exam corrections) performed least well (Figure 1).

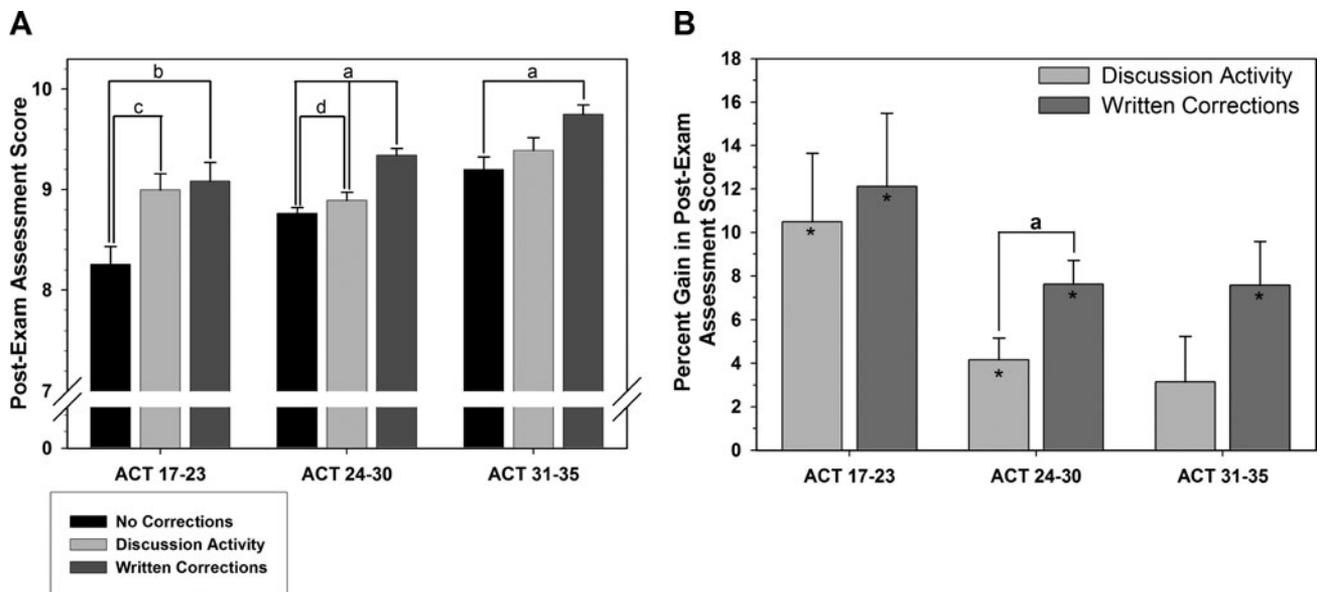
To gain information on the utility of this exercise for students of different academic levels, we sorted the students into groups based on ACT scores. College grade point averages were not available for the majority of the students, as students were primarily first-semester freshman. The average ACT score over the three sections was 27.1 for the 611 students for whom data were available and did not differ significantly between the three sections (mean of 26.8, 27.4, and 27.2 for sections A, B, and C, respectively). Students were grouped into a low (17–23, 15.3% of the students), midrange (24–30, 67% of the students), and high (31–35, 17.7% of the students) ACT group for analysis of results. For all students, a comparison was made of the postexam assessment with no cor-

rections, the discussion activity, and the written corrections. If ACT scores were unavailable for a student, they were not included in the analysis. The performance of 391 students following the discussion activity and 392 students following the written corrections was analyzed with a two-way ANOVA. For all three ACT groups, the highest postexam assessment scores were seen following written corrections (Figure 2A). A two-way ANOVA revealed a significant difference in postexam assessment score with different interventions and different ACT scores ( $p < 0.001$ ). One-way ANOVAs followed by pairwise comparisons using the Holm-Sidak method were performed on each ACT group, revealing that all groups that participated in written corrections demonstrated enhanced performance on the postexam assessment. Only the students in the low and midrange ACT groups appeared to benefit from the discussion activity-based corrections. Only in the midrange ACT group with the largest  $n$  was there a significant difference between the postexam assessment scores of students who participated in written corrections compared with those who participated in discussion activity-based corrections.

For further determination of whether the magnitude of the gains in the different ACT groups differed, the postexam assessment scores for the students participating in the discussion activity and written corrections were calculated as a percent gain over the postexam assessment scores for students who did not formally carry out exam corrections (Figure 2B). Due to variability in the data, the difference in the mean values among the different levels of intervention was not significant utilizing a two-way ANOVA. However, the difference in the mean values based on ACT score were significantly different ( $p = 0.029$ ). A post hoc analysis using the Holm-Sidak procedure revealed that students in the low ACT group significantly increased their performance on the postexam assessments following the discussion activity and written corrections in comparison with either the midrange ( $p = 0.010$ ) or high ( $p = 0.017$ ) ACT groups. To gain more insight into the differences among the different groups, we made comparisons between the various interventions within a single ACT group. With the exception of the students in the high ACT group participating in the discussion activity, in a one-sample  $t$  test, all groups with either the discussion activity or the written corrections performed better on the postexam assessment in comparison with students not performing postexam analyses. As demonstrated with the one-way ANOVA of the postexam assessment scores in Figure 2A, comparison of the gain on postexam assessments following the discussion activity and written corrections with two-sample  $t$  tests demonstrates that only the midrange ACT group performed significantly better following written corrections in comparison with the discussion activity ( $p = 0.021$ ).

### **Writing Assignments Lead to Improved Short- and Long-Term Retention of Information**

Based on assessments administered 2-wk postactivity, students who participated in writing assignments and discussion activities performed significantly better than the lecture-only control groups (Figure 3,  $p = 0.020$  and  $p \leq 0.001$ , respectively). There was no difference in the performance of students writing the essay or participating in an active



**Figure 2.** Students with low ACTs benefit more than other students from exam corrections. (A) Exam-correction improvement data sorted into a low (17–23), midrange (24–30), and high (31–35) ACT group. Data were collected and normalized as described for Figure 1. A two-way ANOVA revealed significant differences in the postexam assessment scores with the different interventions and across the different ACT groups ( $p < 0.001$ ). One-way ANOVA of each ACT group revealed significant differences between interventions for each ACT group ( $p = 0.002$ ,  $p < 0.001$ , and  $p = 0.003$  for the low, midrange, and high ACT groups, respectively). Pairwise comparisons with the Holm-Sidak method revealed differences between no corrections and written corrections for all three ACT groups, between no corrections and a discussion activity for the low and midrange, ACT groups but not for the high ACT group. (a)  $p < 0.001$ ; (b)  $p = 0.001$ ; (c)  $p = 0.003$ ; (d)  $p = 0.013$ . (B) The postexam assessment scores for each student following the written corrections or discussion activity were divided by the postexam assessment score in the absence of any corrections and multiplied by 100 to obtain a percent gain with interventions. A two-way ANOVA revealed that the differences in the mean values based on ACT score were significantly different ( $p = 0.029$ ). A post hoc analysis using the Holm-Sidak procedure revealed that the low ACT group had a significantly increased performance on the postexam assessments following the discussion activity and written corrections in comparison with either the midrange ( $p = 0.011$ ) or high ( $p = 0.017$ ) ACT groups. The average gain on each intervention for each group was compared with 0 with a one-sample  $t$  test to determine whether there was a significant gain over no corrections. The  $p$  value for all groups and all interventions, except the discussion activity in the high ACT group, was  $< 0.002$  (\*). A two-sample  $t$  test was used to compare the average gain with written corrections to the discussion activity for each ACT group. The sample size was 42 for the low ACT group, 279 for the midrange ACT group, and 70 or 71 for the high ACT group, depending on the intervention. (a)  $p = 0.021$ . Data represent mean  $\pm$  SEM.

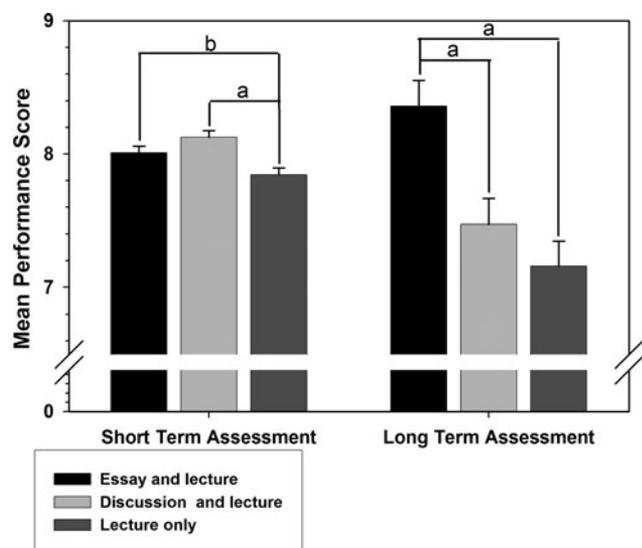
discussion of a topic on this short-term assessment. Interestingly, by the time of the cumulative final exam (long-term assessment), students who had written an essay on a topic performed significantly better than both the students who participated in an active discussion and those who only attended traditional lectures on the topic (Figure 3,  $p \leq 0.001$  for both comparisons using the Holm-Sidak method). There was no significant interaction between the topic and the form of presentation for either the online assessments or the questions on the final exam.

To gain more information on the utility of this exercise for various students, we sorted the students into groups based on ACT scores, as described for the exam-correction assignment. When the students were separated into the different ACT groups, there was no difference in the pattern of the scores on the short-term assessment in the three groups; regardless of ACT score, students performed better after writing an essay or participating in an active discussion when compared with students attending traditional lectures on the topic (unpublished data). On the long-term assessment, however, the pattern of the three interventions was quite different among the three ACT groups (Figure 4). Students with midrange and high ACT scores performed as demonstrated in Figure 3,

with the students participating in the essay performing the best (one-way ANOVA:  $p = 0.005$  for midrange ACTs;  $p = 0.004$  for high ACTs). However, the performance scores for the low ACT students did not differ with the different interventions.

## DISCUSSION

Norris and Phillips (2003) suggested that reading and writing in science are not only tools for information storage and retrieval but are also necessary to promote scientific literacy. In relatively small biology courses for nonmajors, writing assignments have been reported to improve writing skills (Libarkin and Ording, 2012) and, more importantly, critical-thinking skills (Quitadamo and Kurtz, 2007). In addition to writing, the introduction of metacognition to a classroom benefits the students by making them better learners (Tanner, 2012). In this paper, we show that writing interventions that include a metacognitive component in large-enrollment introductory biology courses have significant impacts on the student learning compared with other traditional teaching methods.

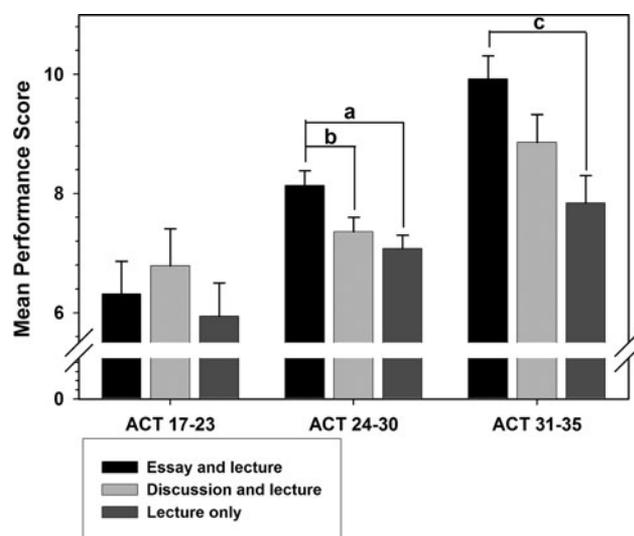


**Figure 3.** Writing assignments increase student short-term and long-term retention. Students participating in interventions (Table 3) were assessed for proficiency 2 wk after the assignment (five-question quiz; short-term retention) and on the cumulative final exam (two exam questions for each topic; long-term retention). In each case, the assessment was scored out of 10 points. Overall, students performed best on the scientific design assessments, so all data were normalized within each topic by multiplying each individual score by the ratio of the average score on the scientific design assessment over the average score on the assessment for the topic as described in *Methods*. The data were analyzed by a two-way ANOVA with the topic and essay/discussion/lecture only as the independent variables ( $p < 0.001$ ); this was followed by the Holm-Sidak method of pairwise comparison. (a)  $p \leq 0.001$ ; (b)  $p = 0.020$ .  $n = 551, 563$ , and 568 for essay and lecture, discussion and lecture, and lecture only, respectively. The normalized data are expressed as mean  $\pm$  SEM.

### Exam Corrections

Some instructors use exams as learning tools. After the exam, students are asked to analyze why they answered questions incorrectly and then research the correct answer. It has been suggested that this process clears up misconceptions and enhances learning. The execution of such an activity can vary from small-group discussions to formal written analysis of incorrect answers. We were interested in whether the practice of formal exam corrections leads to positive quantifiable outcomes in retention of learning and critical thinking and, if so, what approach has the largest impact on student learning and which type of student benefits the most from this exercise.

Exams are generally used by faculty as an assessment tool but are also a valuable teaching tool if they are revisited and reviewed to clarify misunderstandings (Boehm and Gland, 1991; Ley and Young, 2001; Henderson and Harper, 2009). Instructors often encourage students to review their exam answers in order to understand why they answered questions incorrectly. This is a metacognitive process that allows students to identify and correct lingering misunderstandings and provides insight as to how students can improve their studying for the next exam. In the case of higher-order questions that require students to analyze and synthesize information, review of incorrectly answered questions provides an opportunity to practice these skills. Unfortunately, the gen-



**Figure 4.** Effect of writing assignments on long-term retention varies with ACT score for students. Student scores on the final exam questions (two questions for each topic; scored out of 10 points) were normalized as described for Figure 3 and separated based on ACT score for the long-term assessment and analyzed with a one-way ANOVA followed by Holm-Sidak pairwise comparisons. For the low ACT, group:  $n = 58, 61$ , and  $60$  for essay, discussion, and lecture, respectively; for the midrange ACT group:  $n = 358, 348$ , and  $367$ ; and for the high ACT group:  $n = 96, 93$ , and  $95$ . (a)  $p = 0.002$ ; (b)  $p = 0.021$ ; (c)  $p = 0.001$ . The normalized data are expressed as mean  $\pm$  SEM.

eral population of students, especially freshmen, may not understand this process (Harper *et al.*, 2004). Offering exam corrections for credit in introductory biology courses encourages all of the students to look over their exams. For written corrections, the students were required to explain why each wrong answer they gave was incorrect and why the correct answer was correct. This exercise requires students to analyze where their thinking was flawed. With multiple-choice exams, there often may be one best answer with a number of answers that are partially correct. This process of understanding why particular choices are not the best answer is exactly what the students should have been doing during the exam period with multiple-choice questions. A student will not be able to correctly identify the best answer without truly analyzing why the other answers are only partially correct. The written correction assignment requires the students to go through this metacognitive process. With a large class, “grading” of such corrections can be unwieldy. As an alternative, we also assigned “discussion-based corrections” wherein the students discussed the 10 most difficult questions in a group; this was followed by a clicker quiz on the same 10 questions. This task requires the same metacognitive analysis on the students’ part without the added written component.

We demonstrated that students participating in any type of exam-correction activity performed better than those who did not (Figure 1). Interestingly, students writing out corrections performed significantly better than all other groups. Written corrections may be the most effective intervention, because, compared with group discussion of the most difficult questions, this assignment forces students who performed poorly to go over all of their specific misunderstandings of the exam

material in detail. In addition, written corrections are specifically tailored to the students' individual misunderstanding of topics and provide an opportunity to remedy their misconceptions. Written corrections combine the act of writing with the metacognitive processes of analysis and reflection. The data demonstrate that the learning benefits of exam corrections are significant and thus appear to justify the additional time required for grading these corrections or the in-class time dedicated to discussing the previous exam.

William *et al.* (2011) recently reported positive results from their evaluation of exam analysis in their large-enrollment introductory biology class by asking students to write an analysis on a few questions incorrectly answered by the class. The gains reported in that study were topic specific, and only a single assessment question was utilized to determine the performance of the students on a particular topic. Our study indicates that written corrections based on all incorrectly answered questions provide increased learning for students regardless of the subject area in biology.

Sorting the students into groups based on ACT scores revealed that students with low ACT scores benefited more from exam corrections than students with higher ACT scores (Figure 2). In addition, the students with low ACT scores benefited almost equally whether the exam corrections were in a written or discussion format, suggesting that it is the metacognitive component that is most important in their improvement. This group of students is most likely in need of help with how to approach exam questions. These students come to lecture and read the textbook but often lack critical test-taking skills. Writing out the corrections as required for our assignments or discussing the questions in a group should help to train them not only to answer questions in the specific areas but also to perform better on subsequent assessments. The act of learning how to answer a multiple-choice question thoughtfully may account for the higher average assessment scores across all students as the semester progressed.

Students with midrange ACT scores benefited most from writing but also benefited from the discussion activity, whereas the students with the highest ACT scores appeared to benefit only from written corrections. There are a number of possible reasons for the lack of benefit of the discussion activity for the high-end students. These students were already scoring high on the assessments in the absence of "formal" corrections. It is likely that this cohort of students performs exam analyses on their own, in the absence of a course requirement. These high-performing students are most often the ones who want to know why they got an answer incorrect on the exam. This group of students is already proficient at the strategies needed to answer multiple-choice questions. Therefore, they may not benefit as much from the group discussion. Their performance should benefit the most from going over the specific questions they missed, which may or may not overlap with the 10 questions chosen for the discussion activity.

### **Essay Assignments Using Peer Review**

In a large class with only a single teaching assistant for 200–300 students, offering multiple or even one significant writing assignment is not a manageable task. The CPR system offers an opportunity for students not only to write about sci-

ence but also to participate in a type of peer-review process that is intrinsic to how science is evaluated and funded. The assignments we designed required the students to write an analytical-style essay that included more than just rephrasing and summarizing the sources. Students utilized basic information acquired by attending lecture and reading the textbook and applied it in a new context that was driven by reading additional source materials on the topic. This was followed by reviewing other students' essays and scoring these essays on how well they addressed the issues raised in the assignment. Finally, the students were required to reflect upon and analyze their own performance on the essay assignment. The positive benefits of using the CPR program have been reported by a number of groups (Pelaez, 2002; Gunersel *et al.*, 2008; Gunersel and Simpson, 2009; Rourke *et al.*, 2008; Clase *et al.*, 2010), while others have reported no benefit from the program (Walvoord *et al.*, 2007). It is important to look at the design of the assignment and also at the criteria that were used to determine the benefit of the writing assignment to the students. In some studies, a change in the students' perceptions was considered a positive outcome (Clase *et al.*, 2010; Brownell *et al.*, 2013), while scientific writing skills and critical-thinking skills were not specifically tested. Many other studies have specifically analyzed scientific writing skills, with mixed results. Gunersel and Simpson reported an improvement in writing and reviewing skills using the CPR scores as a measure of improvement (Gunersel *et al.*, 2008; Gunersel and Simpson, 2009). Walvoord and colleagues (2007) did not see an increase in writing skills when the essays were scored independently by the director of the university writing center. Pelaez (2002) designed a controlled study using CPR to assign problems in an upper-division physiology course. Students performed better on assessments of content presented only in the CPR assignment when compared with assessments of content presented in lecture. The results of our study are consistent with those of Pelaez in that the improved performance of the students was specific to the topics covered by the writing assignment.

Students who participated in the written essay assignment or small-group discussions in the present study demonstrated in short-term assessments that these two activities provided significant learning gains compared with the control group (Figure 3). In the long-term assessment, however, those students who wrote on a topic performed even better than those students who participated in an active group discussion (Figure 3). Ultimately, performance on any assessment that requires critical thinking involves a combination of recall and analytical ability. Students participating in the essay and discussion activities may have performed well in the short term because their activity included application of the content presented in lecture. For the long-term assessment, students writing an essay on the topic performed significantly better than the other groups. This may be due to the fact that, in addition to the analytical and metacognitive aspects of their assignment, they had the added benefit of enhanced recall due to the writing process (Weinstein and Mayer, 1983). Overall, students performed differently on the assessments for each of the three topics, with the best quantitative performance occurring on scientific design and the worst on energy and macromolecules. Two factors are likely to have contributed to these differences: 1) The order of assigned topics throughout the semester was energy and macromolecules,

CO<sub>2</sub> and ocean acidification, and, finally, scientific design. It is likely that students gained proficiency in assessment performance as the semester progressed and they became familiar with the instructor and question styles. 2) The later topics also received more class time, because topics such as acids and bases and scientific design were brought up in a number of different contexts throughout the semester. For this reason, the raw scores were normalized to the performance on scientific design to account for this difference (see *Methods*).

Our experimental approach did not distinguish whether the enhanced learning observed with the students who participated in the writing assignment was due to the active writing itself or to increased “time on task” for these students. Students involved in the writing assignment are believed to have spent more time on the assigned topic than did students in the other two groups and were also forced to work with the material at a higher cognitive level by synthesizing information and evaluating their peers and themselves. In contrast, students participating in the group discussion or traditional lecture only spent a limited amount of formal contact time on these topics. It is unknown whether students in these two groups engaged in other activities beyond the formal material/presentations that were provided in the lecture and/or discussion. It is assumed that, in general, increased time spent in effective studying will result in better comprehension of the material. It would be interesting to know the extent to which the improvement observed with this assignment is related to the higher-order functions involved with the writing/reviewing versus the extended time on task. Further studies would be appropriate to control for time on task.

It is interesting to note that the benefits of the writing assignment appear limited to students with midrange or high ACT scores (Figure 4). Students with ACT scores of 23 or lower did not significantly raise their assessment scores following a writing assignment or active discussion. It may be that this population of students struggled more with the basic content, which impeded their ability to raise their analysis to a higher level. Another possibility is that these students may have been too far behind to benefit from these interventions. This is in direct contrast to the effects of exam corrections on subsequent assessments in this group of students, wherein the corrections provided a means for students to improve in basic knowledge and correct misconceptions. This deserves further investigation.

In conclusion, offering opportunities to write with a metacognitive component is beneficial to students’ abilities to answer higher-order questions in subsequent assessments. We identified differences in students based on their ACT scores. In general, the benefit of various interventions was different for students with low ACT scores in comparison with those with midrange or high ACT scores. Writing an analytical essay was not particularly beneficial to the low ACT students, suggesting they may need more guidance to benefit from this type of assignment. This is in contrast to the benefits of either written or discussion-based exam corrections, which appear to be most beneficial to the low ACT group of students. We have demonstrated that essay writing with the CPR program and formal exam corrections can both be designed in such a way as to be feasible in a large introductory class with minimal help from teaching assistants. It is envisioned that these writing interventions can be incorporated into any STEM course, regardless of level or size, with

similarly significant improvements in learning outcomes and minimal additional work for the instructor.

## ACKNOWLEDGMENTS

This work was supported by the Way-Klingler College of Arts and Sciences 2011 Way-Klingler Teaching Enhancement Award granted to T.J.E., M.M., and M.St.M. We acknowledge the support of A. Riley and L. MacBride, Institutional Research at Marquette University, in data access, handling, and analysis. We thank Lisa Petrella, Marquette University, and Shannon Colton and Gina Vogt, program directors of the Center for Biomolecular Modeling at the Milwaukee School of Engineering, for categorizing the assessment questions. The authors also acknowledge the teaching and publishing community for many interactive discussions on teaching at a variety of workshops, with a particular mention of the National Academies Summer Institute on Undergraduate Education in Biology ([www.academiessummerinstitute.org](http://www.academiessummerinstitute.org)), which T.J.E. and M.M. attended in 2006.

## REFERENCES

- Ainley M, Hidi S, Berndorff D (2002). Interest, learning, and the psychological processes that mediate their relationship. *J Educ Psychol* 94, 545–561.
- Bangert-Drowns RL, Hurley MM, Wilkinson B (2004). The effects of school-based writing-to-learn interventions on academic achievement: a meta-analysis. *Rev Educ Res* 74, 29–58.
- Bloom BS (1984). *Taxonomy of Educational Objectives, Book 1: Cognitive Domain*, Boston, MA: Addison Wesley.
- Boehm R, Gland JL (1991). Using exams to teach chemistry more effectively. *J Chem Educ* 68, 455.
- Brownell SE, Price JV, Steinman L (2013). A writing-intensive course improves biology undergraduates’ perception and confidence of their abilities to read scientific literature and communicate science. *Adv Physiol Educ* 37, 70–79.
- Clase KL, Gundlach E, Pelaez NJ (2010). Calibrated peer review for computer-assisted learning of biological research competencies. *Biochem Mol Biol Educ* 38, 290–295.
- Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom’s taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7, 368–381.
- Cutler WB, Friedmann E, McCoy NL (1998). Pheromonal influences on sociosexual behavior in men. *Arch Sex Behav* 27, 1–13.
- Durst RK, Newell GE (1989). The use of function: James Britton’s category system and research on writing. *Rev Educ Res* 59, 375–394.
- Gourgey AF (2002). Metacognition in basic skills instructions. In: *Metacognition in Learning and Instruction. Theory, Research and Practice*, ed. HJ Hartman, Norwell, MA: Kluwer Academic.
- Gunersel A, Simpson N (2009). Improvement in writing and reviewing skills with Calibrated Peer Review™. *Int J Scholarship Teach Learn* 3, article 15.
- Gunersel AB, Simpson NJ, Aufderheide KJ, Wang L (2008). Effectiveness of Calibrated Peer Review for improving writing and critical thinking skills in biology undergraduate students. *J Scholarship Teach Learn* 8(2), 25–37.
- Harper KA, Brown RW, Finnerty M (2004). A treatment for post-exam syndrome. Paper presented at the 2004 AAPT Winter Meeting, held 26 January 2004 in Miami Beach, FL.
- Heidemann M, Urquhart G (2005). A Can of Bull? Do Energy Drinks Really Provide a Source of Energy? National Center for Case Study Teaching in Science. [http://sciencecases.lib.buffalo.edu/cs/files/energy\\_drinks.pdf](http://sciencecases.lib.buffalo.edu/cs/files/energy_drinks.pdf).

- Henderson C, Harper KA (2009). Quiz corrections: improving learning by encouraging students to reflect on their mistakes. *Phys Teach* 47, 581.
- Hoegh-Guldberg O *et al.* (2007). Coral reefs under rapid climate change and ocean acidification. *Science* 318, 1737–1742.
- Holt S (2002). Love Potion #10, National Center for Case Study Teaching in Science. [http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case\\_id=173&id=173](http://sciencecases.lib.buffalo.edu/cs/collection/detail.asp?case_id=173&id=173).
- Ley K, Young DB (2001). Instructional principles for self-regulation. *Educ Tech Res Dev* 49, 93–103.
- Libarkin J, Ording G (2012). The utility of writing assignments in undergraduate bioscience. *CBE Life Sci Educ* 11, 39–46.
- Norris SP, Phillips LM (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Educ* 87, 224–240.
- Paris SG, Paris AH (2001). Classroom applications of research on self-regulated learning. *Educ Psychol* 36, 89–101.
- Pelaez N (2002). Problem-based writing with peer review improves academic performance in physiology. *Adv Physiol Educ* 26, 174–184.
- Quitadamo I, Kurtz M (2007). Learning to improve: using writing to increase critical thinking performance in general education biology. *CBE Life Sci Educ* 6, 140–154.
- Reynolds JA, Thaiss C, Katkin W, Thompson RJ (2012). Writing-to-learn in undergraduate science education: a community-based, conceptually driven approach. *CBE Life Sci Educ* 11, 17–25.
- Roediger HL, III, Butler AC (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn Sci* 15, 20–27.
- Roediger HL, III, Karpicke JD (2006). The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci* 1, 181–210.
- Rourke AJ, Mendolsohn J, Coleman K, Allen B (2008). Did I mention it's anonymous? The triumphs and pitfalls of online peer review. In: Hello! Where Are You in the Landscape of Educational Technology? Proceedings ascilite Melbourne 2008, 830–840. [www.ascilite.org.au/conferences/melbourne08/procs/rourke.pdf](http://www.ascilite.org.au/conferences/melbourne08/procs/rourke.pdf).
- Russell AA, Chapman OL, Wegner PA (1998). Molecular science: network-deliverable curricula. *J Chem Educ* 75, 578–579.
- Schraw G (2002). Promoting general metacognitive awareness. In: *Metacognition in Learning and Instruction. Theory, Research and Practice*, ed. HJ Hartman, Norwell, MA: Kluwer Academic.
- Tanner KD (2012). Promoting student metacognition. *CBE Life Sci Educ* 11, 113–120.
- Walvoord ME, Hoefnagels MH, Gaffin DD, Chumchal MM, Long DA (2007). An analysis of Calibrated Peer Review (CPR) in a science lecture classroom. *J Coll Sci Teach* 37, 66–73.
- Weinstein CE, Mayer RE (1983). The teaching of learning strategies. In: *Handbook of Research on Teaching*, ed. MC Wittrock, New York: MacMillan, 315–327.
- William AE, Aguilar-Roca NM, Tsai M, Wong M, Beaupre MM, O'Dowd DK (2011). Assessment of learning gains associated with independent exam analysis in introductory biology. *CBE Life Sci Educ* 10, 346–356.
- Yerushalmi E, Henderson C, Heller K, Heller P, Kuo V (2007). Physics faculty beliefs and values about the teaching and learning of problem solving. I. Mapping the common core. *Phys Rev ST Phys Educ Res* 3, 020109.
- Zumbrunn S, Tadlock J, Roberts ED (2011). *Encouraging Self-Regulated Learning in the Classroom: A Review of the Literature*, Richmond: Metropolitan Education Research Consortium, Virginia Commonwealth University, 1–28.