Title
Practice makes pretty good: assessment of primary literature reading abilities across multiple large-enrollment biology laboratory courses.

Journal
CBE life sciences education, 13(4)

ISSN
1931-7913

Authors
Sato, Brian K
Kadandale, Pavan
He, Wenliang
et al.

Publication Date
2014

DOI
10.1187/cbe.14-02-0025

Peer reviewed

CBE—Life Sciences Education
Vol. 13, 677–686, Winter 2014

Article

Practice Makes Pretty Good: Assessment of Primary Literature Reading Abilities across Multiple Large-Enrollment Biology Laboratory Courses

Brian K. Sato,*† Pavan Kadandale,*† Wenliang He,‡ Paige M. N. Murata,* Yama Latif,* and Mark Warschauer‡

*Department of Molecular Biology and Biochemistry and ‡Department of Education, University of California, Irvine, Irvine, CA 92697

Submitted February 10, 2014; Revised August 27, 2014; Accepted August 27, 2014
Monitoring Editor: James Hewlett

Primary literature is essential for scientific communication and is commonly utilized in undergraduate biology education. Despite this, there is often little time spent *training* our students how to critically analyze a paper. To address this, we introduced a primary literature module in multiple upper-division laboratory courses. In this module, instructors conduct classroom discussions that dissect a paper as researchers do. While previous work has identified classroom interventions that improve primary literature comprehension within a single course, our goal was to determine whether including a scientific paper module in our classes could produce long-term benefits. On the basis of performance in an assessment exam, we found that our module resulted in longitudinal gains, including increased comprehension and critical-thinking abilities in subsequent lab courses. These learning gains were specific to courses utilizing our module, as no longitudinal gains were seen in students who had taken other upper-division labs that lacked extensive primary literature discussion. In addition, we assessed whether performance on our assessment correlated with a variety of factors, including grade point average, course performance, research background, and self-reported confidence in understanding of the article. Furthermore, all of the study conclusions are independent of biology disciplines, as we observe similar trends within each course.

INTRODUCTION

Scientific discovery is predicated on our ability to build upon the work of our peers and previous generations of researchers. This information is transmitted in a number of ways, including presentations at scientific meetings, personal communications, and, most importantly, published literature. Primary research articles allow data to be shared across the world instantaneously, providing a nearly endless library of information. Not only are they vital for professional scientists, but papers are also valuable tools for educators and undergraduates. The use of scientific papers in the classroom has been shown to increase student learning and interest in the course material (DebBurman, 2002; Luckie *et al.*, 2004; Kozeracki *et al.*, 2006; Hoskins *et al.*, 2007). In addition, the skills involved in being able to read and critically analyze papers are applicable across subject areas and are not specific to biology. Despite the value of primary literature for budding scientists, from freshmen in an introductory biology course to graduate students, most curricula lack any formal training on how to critically read scientific papers.

To fill this void, techniques have been established to increase students' abilities to read, comprehend, and analyze primary literature. A few examples include the CREATE method, Figure Facts, and Research Deconstruction (Clark *et al.*, 2009; Hoskins *et al.*, 2007, 2011; Gottesman and Hoskins,

DOI: 10.1187/cbe.14-02-0025
†These authors contributed equally to this work.
Address correspondence to: Brian K. Sato (bsato@uci.edu).

2013; Round and Campbell, 2013). While these methods are valuable additions to science, technology, engineering, and mathematics (STEM) learning, not much is known about whether the analytical skills gained from these practices are evident in later courses. Additionally, it is unknown whether the measured performance gains are specific for certain biology fields. At the University of California, Irvine, more than 90% of the students participating in this study (mostly fourth- and fifth-year students who were enrolled in biology lab classes) reported having been exposed to primary literature in previous courses. Yet from personal communication with students and in classroom discussions, it is clear that the average student is lacking in this skill.

To address this problem, we established a common primary literature module running through three of our large-enrollment upper-division laboratory courses, including microbiology lab, molecular biology lab, and biochemistry lab. The hallmark of this module was an instructor-led modeling of how a researcher approaches a scientific article (see *Methods*). Students were tasked with reading three research papers throughout the course, the last of which was used in an assessment of their comprehension and critical-thinking abilities in an examination known as the "paper quiz." All three papers examined each quarter were relevant to the topics of the specific lab course, and the paper quiz articles were unique for each course and quarter.

The goal of this study was to determine whether this module would result in long-term improvements in students' abilities to comprehend and critically analyze data from primary sources, irrespective of the specific discipline of biology such data were drawn from. Such a longitudinal gain, if present, would be significant, as our goal as educators should be to provide our students with skills that are transferable beyond the immediate classroom in which they were acquired. We evaluated the long-term effects of the module by comparing paper quiz performance between "first-time" students (students taking one of the labs included in the study for the first time) and "returner" students (students who have previously taken one of the labs included in the study). Our measure of student comprehension and critical-thinking skills was performance on questions classified by Bloom's taxonomy (Krathwohl, 2002; Crowe *et al.*, 2008). Student comprehension of primary literature was gauged based on performance on lower-level Bloom's questions (Bloom's 1 and 2), while critical analysis was measured by performance on Bloom's 3–6 questions (application, analysis, evaluation, synthesis). Critical thinking encompasses a variety of skills researchers use when approaching a scientific paper. This includes the ability to interpret figures, use data to predict outcomes, design experiments based on author conclusions, and assess the scientific validity of arguments made in the paper. Higher-order Bloom's questions can assess these skills and are commonly used across multiple disciplines to assess student critical-thinking abilities (Fuller, 1997; Athanassiou *et al.*, 2003; Bissell and Lemons, 2006; Crowe *et al.*, 2008; Stanger-Hall, 2012). If our module resulted in increased comprehension and critical-thinking abilities, we would expect to see higher performance on questions of various Bloom's levels for returner students relative to first-timers.

For study purposes, student performance on paper quizzes was analyzed based on a variety of parameters, including educational background, study method utilized, and student confidence, among others. While we were able to identify a number of variables that impacted paper quiz scores, one of the strongest conclusions was that exposure to this module increased paper-reading comprehension and analysis in subsequent lab courses (as measured by higher scores on questions of various Bloom's levels), demonstrating a longitudinal benefit rarely documented in STEM education studies. In addition, student reading of primary literature was independent of the specific lab course, as similar trends were identified in each of the three biology labs examined. This implies that dissecting scientific articles as a core component of one or multiple classes increases student learning, regardless of the presence of specific learning techniques to accompany these discussions.

METHODS

Primary Literature Module Description

Student data were collected from three courses, Bio Sci M114L (biochemistry lab), Bio Sci M116L (molecular biology lab), and Bio Sci M118L (microbiology lab) during the 2012–2013 academic year, encompassing nine distinct courses. Over the three quarters of the study, enrollments ranged from 40 to 160 students per class, totaling roughly 900 students. The primary literature reading module was only one part of each course; students were also required to take weekly quizzes, write lab reports based on their experiments, and peer review other students' writing samples. In the molecular biology and biochemistry labs, students were also given a cumulative final exam. Throughout the study period, one instructor (B.S.) taught both the microbiology and biochemistry labs, while a second instructor (P.K.) taught the molecular biology labs.

In the module, three papers were presented during the quarter. For paper 1, students were given an assignment (Study Method 1) in which they were required to answer the following questions regarding selected figures in one to two sentences:

1. Why was the experiment performed?
2. How was the experiment performed?
3. What were the results obtained?
4. What conclusions were made?

For paper 2, students were given an assignment (Study Method 2) to write summary paragraphs (of unspecified length) regarding selected figures. Both of these papers were dissected in lecture the following week, with the instructor guiding the discussion. Class discussions included how to determine the main purpose of the paper, the purpose of each figure, the experimental design for each figure, conclusions derived from the data, how those conclusions are related to the overall goal of the paper, and potential future directions (further description is presented in the Supplemental Material).

Paper 3 was assigned to the class a week before the paper quiz. Students were free to study in any manner, although they were encouraged to use method 1 or 2. The assigned article for each paper quiz was unique, and the quiz consisted of questions ranging from Bloom's level 1 to level 6 (Krathwohl, 2002; Crowe *et al.*, 2008). In addition, each quiz included ungraded questions for students to self-report their independent research background, the study method utilized to prepare for the paper quiz, and their confidence in their understanding of the paper as measured on a Likert scale. Example paper quizzes are provided as Supplemental Material.

This study was performed with approval from the University of California, Irvine, Institutional Review Board (HS# 2012–9026).

Statistical Analysis of Data

Because this study was observational in nature, the multiple-regression technique was used to control for variables that were correlated with one another. For each of the nine courses included in the study, the final paper and therefore the paper quiz were not the same to minimize any potential advantages students might acquire by referencing previous exams. However, the structure of the quizzes was matched to ensure meaningful comparison within and between quarters. The weights we used to compute the composite paper quiz score were the average proportion of questions across nine quizzes on each Bloom's level. They were 32.8% (Bloom's 1 and 2), 19.4% (Bloom's 3), 21.4% (Bloom's 4), 17.8% (Bloom's 5), and 8.6% (Bloom's 6). Paper quiz scores accounted for 10% of the final grade in each class, which was directly converted to a numerical value from "F" = 0 to "A+" = 12. Confidence was measured on a Likert scale from 1 to 5, with 5 being the most confident. This ordinal variable was treated as a continuous variable in the regression analysis.

In the multiple-regression analyses, all dependent variables (scores on questions of each Bloom's level, composite paper quiz score, final course grade, and confidence) were continuous variables. Most independent variables, however, were categorical and therefore were dummy coded on each component level. Specially, returner students were compared with first-time students; study methods (1 or 2) and the degree to which the students utilized them (on all figures or on some figures) were compared with study method 1 utilized for all figures; students with medical and no research experience were compared with students with bench research experience; and molecular biology lab and microbiology lab were compared with biochemistry lab. Grade point average (GPA) and confidence were entered as continuous predictors.

To be noted, all multiple-regression models in our study were class fixed-effects models. As stated previously, all papers and quizzes were unique, and students from each class might also differ in important ways, for example, in their average GPA. Therefore, students from some classes could score, on average, significantly higher or lower than students from other classes. Including each class as a dummy variable accommodated the effects due to differences in paper quizzes or students' characteristics. As a result, our class fixed-effects models only compared students with others within the same class, but not between classes.

RESULTS

Student Quiz Performance Overall versus Questions of a Specific Bloom's Level

Rather than measuring only overall paper quiz scores, we examined student performance on questions from all Bloom's levels. The questions on each paper quiz were "Bloomed" by multiple faculty members familiar with the classification system. Not surprisingly, students scored highest on Bloom's level 1 and 2 questions, which required memorization or simple restatements of given information (Figure 1A and Table 1). While higher education instructors like to believe that a strong emphasis is being placed on critical thinking, this is often not reflected in classroom assessments, as the exams that students often encounter consist mostly of fact-recall questions (Momsen *et al.*, 2010). Student performance was significantly lower on Bloom's level 3–6 questions compared with Bloom's 1 and 2 questions ($p < 0.001$). These higher-order questions measure students' abilities to predict an outcome, interpret data, develop novel information from a variety of sources, and critique the experimental design, respectively (Crowe *et al.*, 2008), all key aspects of paper reading. To determine whether these conclusions were discipline specific, we performed the same analysis within each course. For each course, we saw the same trends as with the overall data, with performance on Bloom's 1 and 2 questions significantly higher than performance on the questions requiring higher-order thinking (Figure 1B, Table 2, and Supplemental Table S1; $p < 0.001$).

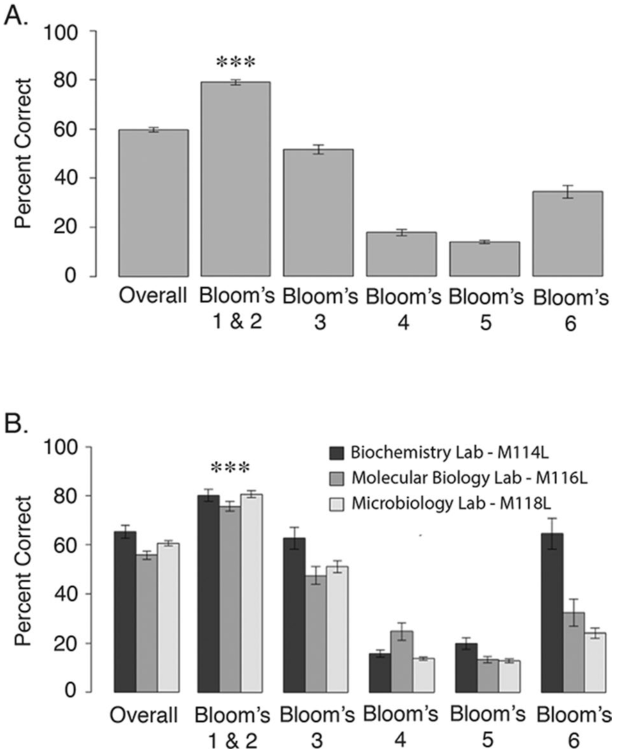

Figure 1. Overall student performance on paper quiz and performance on questions of specific Bloom's levels. (A) Raw scores from nine paper quizzes (three quarters each of microbiology lab, molecular biology lab, and biochemistry lab) are shown. Questions were then broken down by Bloom's level, and performance on each level is indicated. Performance on questions of Bloom's 3–6 was significantly lower than Bloom's 1 and 2 questions (***, $p < 0.001$). (B) Student performance (overall and by Bloom's level) in each course is shown. Performance on Bloom's 3–6 was significantly lower than Bloom's 1 and 2 questions (***, $p < 0.001$) in each of the three courses.

Table 1. Descriptive statistics of key variables[a]

	Count	Mean	SD	Minimum	Maximum
Final grade	892	7.73	2.35	0.00	12.00
Paper quiz score	889	44.28	11.72	14.06	86.49
Bloom's levels 1–2	892	78.94	15.86	0.00	100.00
Bloom's level 3	892	51.50	28.34	0.00	100.00
Bloom's level 4	890	17.72	19.93	0.00	100.00
Bloom's level 5	890	13.86	10.76	0.00	50.00
Bloom's level 6	626	34.25	33.05	0.00	100.00
Confidence in quiz	881	3.57	0.80	1.00	5.00
College GPA	885	3.23	0.42	1.94	4.00
Returner	589	0.18	0.38	0.00	1.00
Method 1 (all figures)	837	0.32	0.47	0.00	1.00
Method 2 (all figures)	837	0.10	0.30	0.00	1.00
Method 1 (some figures)	837	0.22	0.41	0.00	1.00
Method 2 (some figures)	837	0.07	0.25	0.00	1.00
Other method	837	0.29	0.46	0.00	1.00
Bench research	572	0.58	0.49	0.00	1.00
Medical research	572	0.11	0.32	0.00	1.00
No research	572	0.31	0.46	0.00	1.00

[a]The values in the table show the actual minimum and maximum values, not the possible range. Final grade was directly converted from the final course letter grade, coded from 0 ("F") to 12 ("A+"). Confidence in quiz understanding was measured on a five-point (1 for least confident, 5 for most confident) Likert scale. GPA was on a 4.0 scale. Variables including returner status, study methods (method 1 or 2, use of method on all figures or partial usage, or other method), and prior research experience were dummy coded as 0s and 1s. In these cases, the mean value refers to the fraction of the population in each category.

To present the remaining statistical analyses accurately, we modified how the overall quiz scores were represented. Because the composition of exams between the lab courses differed slightly (total points, percentage of questions of each Bloom's level, value of each question), and total quiz scores are a function of the individual questions making up the test, it was necessary to standardize the exams to correctly compare the overall scores between classes. To accomplish this, we generated a composite paper quiz comprising the average points associated with questions from each Bloom's level on all nine paper quizzes. The composite quiz contained 33% Bloom's level 1 and 2 questions, 19% level 3, 21% level 4, 18% level 5, and 9% level 6. The mean scores from each class were then normalized to this composite quiz (Table 1). The overall average in the paper quiz was between 40 and 50%, and we attribute this to the high difficulty level of the quizzes (based on feedback from students in the course, teaching assistants for the course, and faculty who were not involved in the course).

Comparison of Performance between First-Time and Returner Students

As our analysis covered three distinct lab courses, a subset of students chose to take more than one of these courses during the 2012–2013 academic year. For example, one could take microbiology lab Fall quarter and then molecular biology lab Winter quarter. While self-selecting, these returner students were statistically identical to their first-time peers in all measured metrics, including student GPA, course grade in the previous lab course, research experience, and study method utilized on the quiz (unpublished data). In addition, returner students did not have higher averages on the paper quiz compared with their peers in their first lab course (before they were returners). We examined the paper quiz scores of returner students in the Winter and Spring quarters and found their overall scores were significantly higher than first-time students, as was their performance on questions of multiple Bloom's levels (Figure 2A and Tables 2 and S1).

We next determined whether the improvement was due to enrollment specifically in our labs containing the primary literature module or whether any upper-division lab course would produce the observed learning gains. We compared the performance of students who had taken one of two other heavily enrolled laboratory courses before taking our labs ("experienced" students) versus those for whom the microbiology, molecular biology, or biochemistry labs were their first lab experience ("novice"). Experienced and novice students were identical by all statistical measures (unpublished data). Despite the fact that the other two lab courses incorporated primary literature into the curriculum, although not as a core component, the experienced students did not see elevated paper quiz scores compared with the novices. In fact, experienced students exhibited a statistically significant decrease in overall performance and Bloom's level 6 questions (Figure 2B). Thus, the longitudinal learning gains seen by the returner students illustrate that being exposed to a course that involves in-depth discussion of scientific papers is sufficient to increase a student's ability to read and analyze primary literature.

Correlation between Study Method and Student Performance on Paper Quiz

Students were presented with the assigned paper a week in advance of the paper quiz. On the test, we asked students to self-report the method by which they prepared. While they were encouraged to use methods 1 and 2 described above, students were free to study in any manner of their choosing. Students reported whether they studied using method

Table 2. Multiple regression analysis in terms of standardized beta coefficients[a]

	(1) Bloom's levels 1 and 2	(2) Bloom's level 3	(3) Bloom's level 4	(4) Bloom's level 5	(5) Bloom's level 6	(6) Paper quiz score	(7) Final grade	(8) Confidence in paper quiz
Repeater	0.092*	0.096*	0.051	0.087*	0.051	0.131***	0.067*	0.057
	(0.013)	(0.012)	(0.158)	(0.021)	(0.102)	(0.000)	(0.028)	(0.176)
Method 2 (all figures)	–0.107**	–0.083*	–0.041	0.011	–0.022	–0.104**	–0.045	–0.087*
	(0.006)	(0.036)	(0.279)	(0.774)	(0.505)	(0.003)	(0.158)	(0.047)
Method 1 (some figures)	–0.040	–0.079	–0.003	0.071	0.055	–0.029	–0.043	–0.034
	(0.319)	(0.056)	(0.936)	(0.081)	(0.103)	(0.419)	(0.192)	(0.449)
Method 2 (some figures)	–0.036	–0.017	0.012	0.024	0.040	–0.005	–0.025	–0.072
	(0.345)	(0.662)	(0.750)	(0.533)	(0.207)	(0.887)	(0.416)	(0.089)
Other method	0.010	0.010	0.036	0.038	–0.046	0.020	–0.013	–0.093*
	(0.808)	(0.815)	(0.372)	(0.370)	(0.182)	(0.580)	(0.697)	(0.045)
Medical research	–0.062	–0.049	–0.079*	–0.022	–0.056	–0.099**	–0.025	–0.028
	(0.100)	(0.205)	(0.034)	(0.562)	(0.075)	(0.003)	(0.410)	(0.505)
No research	0.001	0.022	0.030	–0.028	0.009	0.020	0.062	–0.176***
	(0.979)	(0.594)	(0.452)	(0.499)	(0.792)	(0.571)	(0.063)	(0.000)
GPA	0.410***	0.265***	0.234***	0.299***	0.160***	0.483***	0.724***	0.067
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.132)
Confidence	0.105**	0.119**	0.071	0.041	0.134***	0.166***	0.071*	
	(0.007)	(0.003)	(0.061)	(0.293)	(0.000)	(0.000)	(0.026)	
Molecular biology lab	–0.008	0.195*	0.652***	–0.448***	0.343***	0.364***	0.217***	–0.184*
	(0.921)	(0.015)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.037)
Microbiology lab	–0.165*	0.194*	0.134	–0.436***	0.019	0.001	–0.015	–0.184*
	(0.037)	(0.017)	(0.085)	(0.000)	(0.770)	(0.994)	(0.824)	(0.040)
Observations	541	541	541	541	541	541	541	541
R^2	0.313	0.278	0.335	0.290	0.514	0.460	0.533	0.121

[a]The coefficients in the table are standardized beta coefficients. As a result, they are readily comparable in magnitude across rows and columns, and the coefficients must be interpreted in terms of SDs. Method 1 (all figures) was used as baseline for comparing against other methods. Similarly, students with bench research experience and those enrolled into the biochemistry lab served as the corresponding baselines. Coefficients for dummy variables of each specific class were omitted for simplicity; they represented the fixed effects of individual classes, which were not of interest in this study.

p Values in parentheses: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

1 or 2 for the entire paper, using method 1 or 2 on some of the figures in the paper (partial usage), or in their own way. Roughly 30% of students prepared with method 1 for the entire paper, 20% with method 2 for the entire paper, 5–10% using method 1 or 2 for some figures, and 30% studied with their own method. Completely using study method 1 or 2 (compared with partial use or no use) did not produce a statistically significant increase in performance (Figure 2C and Tables 2 and S1). We did observe a decrease in performance when students reported partially using either study method though. This drop in performance may be due to the fact that a complete understanding of the paper was required to excel on the paper quiz. It may also be a sign of students half-heartedly attempting to understand the article. Thus, the longitudinal gains cannot be attributed to the methods we encouraged the students to prepare with and instead imply that being introduced to the process of dissecting a paper contributed to the increased performance.

Correlation between Academic Ability, Research Experience, and Student Performance on the Paper Quiz

Nearly two-thirds of biology undergraduates at University of California, Irvine, have conducted independent research during their college career, including both basic science research (bench research) and medical/clinical research. These activities involve participation in the scientific method, and we would imagine that those with bench research experience have been previously exposed to the molecular-based primary literature that was the focus in these courses. To determine whether there is a link between research experience and primary literature reading ability, we asked the students to self-report their research background with one of three options; whether they had bench research experience, medical research experience, or no research experience at all. By analyzing the paper quiz scores in the context of research training, we found that those without research experience performed similarly to those with experience, but students with clinical experience demonstrated a decrease in overall quiz scores compared with those with bench research experience (Figure 3A and Tables 2 and S1). There were no disparities between students with bench or medical research by all measured parameters (unpublished data), and thus we speculated that the distinctions between these research experiences had a role in the differing paper quiz performances.

We also examined the relationship between performance on the paper quiz and overall performance in the class. While the paper quiz only accounted for 10% of the course grade, there was a very strong correlation between

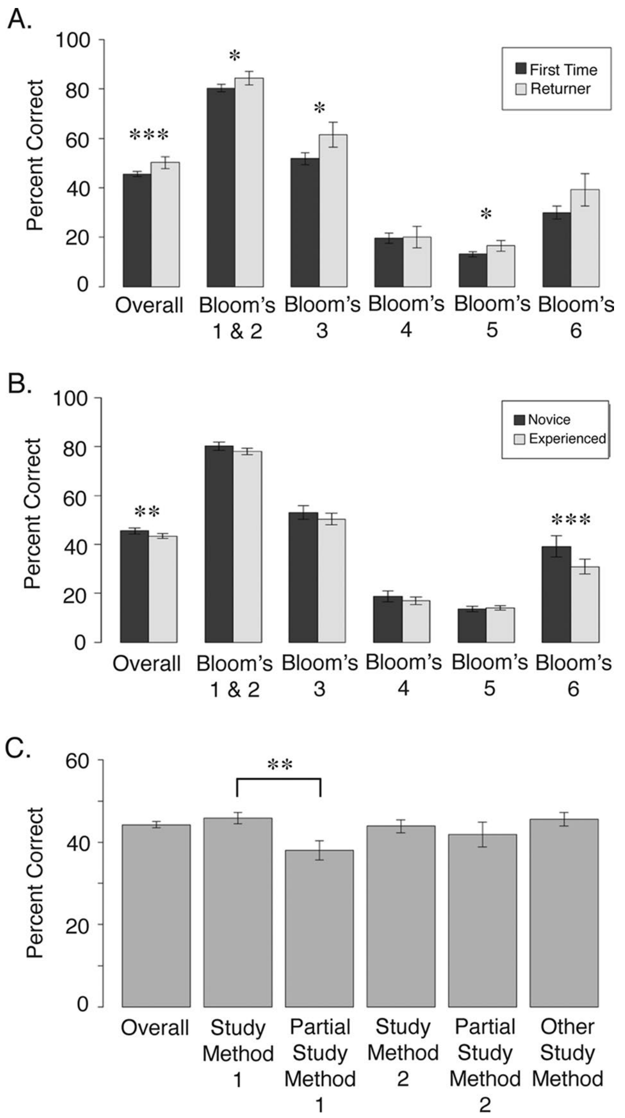

Figure 2. Student performance on paper quiz based on returner statues and preparation method. (A) Overall student performance and performance on questions of each Bloom's level. Returners are students who had taken one of the lab courses involved in the study in a previous quarter and enrolled in a different lab at a later time. The performance in the later quarter was measured for returners. First-time and returner students were identical in terms of average GPA. Returner students performed identically to their nonreturner peers on the paper quiz in their first lab (before they became returners). (B) Overall student performance and performance on questions of each Bloom's level. Experienced students are those who had taken an upper-division lab course that did not include the primary literature module utilized in the study classes. Novice students are those taking one of the courses included in the study as their first biology lab. (C) Overall student scores on the paper quiz are presented based on student study method. Students self-reported how they prepared for the quiz. Method 1 required answering four questions regarding each figure, while method 2 involved writing summary paragraphs for each figure. Partial use of the study method indicates that the given method was not used for all figures in each tested paper. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

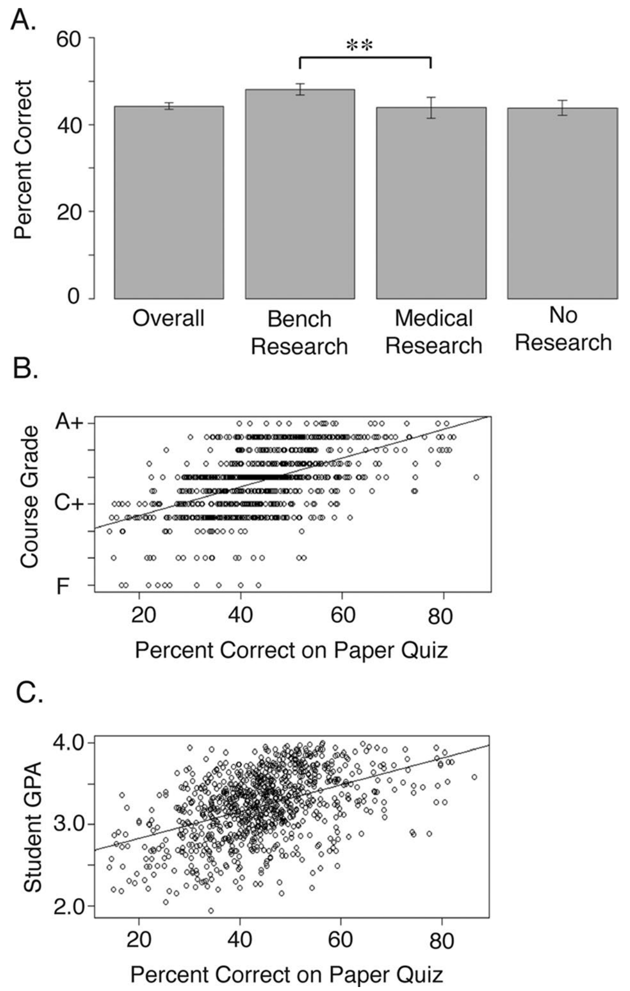

Figure 3. Overall student performance on the paper quiz and relationship to research experience, class grade, and GPA. (A) Overall student scores on the paper quiz are presented based on student self-reported research experience. (B) Linear-regression analysis illustrating the relationship between course grade and overall paper quiz score. (C) Linear-regression analysis illustrating the relationship between student GPA and overall paper quiz score. **, $p < 0.01$.

the two (Figure 3B and Tables 2 and S1). In addition, student GPA was a robust predictor of paper quiz performance (Figure 3C and Tables 2 and S1). One could imagine that the skills required to read primary literature, including hypothesis construction, data analysis, and experimental design, are essential for all aspects of a laboratory course, thus allowing students who can read papers to excel at all course-related tasks. On the other hand, it may be that the most capable students can handle any academic challenge presented to them. To distinguish between these two possibilities, we examined the final grade for returner students. As mentioned before, in their first lab course taken, these future returner students performed identically to their peers (no significant difference in both paper quiz score [$p = 0.64$] and overall course grade [$p = 0.32$]), yet scored higher on the paper quiz when enrolling in another of our lab courses (Figure 2A). By multiple-regression analysis, these students also earned significantly higher grades in the second lab ($p = 0.028$), implying that the increased training for reading primary literature assisted with the other course assignments, including lab reports and exams requiring data analysis (Tables 2 and

S1). On the other hand, experienced students (those who had taken other lab courses previously) and novice students (those taking one of our labs as their first lab experience) earned similar grades in our courses ($p = 0.060$). This implies that, while prior experience in a lab course is not sufficient to raise future performance on lab-related activities, increased primary literature comprehension and critical-thinking abilities may do so.

Correlation between Student Confidence and Performance on the Paper Quiz

In addition to quiz performance, we asked students to rate their agreement with the statement "I am very confident that I understand the paper being tested" on a five-point Likert scale ranging from 5 = strongly agree to 1 = strongly disagree. There are a number of published studies regarding student confidence and its relationship with student background and performance in the classroom (Livengood, 1992; Abouserie, 1995; Morgan and Cleave-Hogg, 2002; Gore, 2006; Valdez *et al.*, 2006). Among these studies, results have varied in regard to whether more-confident students exhibited increased performance on the included assessments.

We first examined whether student confidence was an accurate predictor of overall performance on the paper quiz and found that to be the case, as students who were more confident earned a higher score on the quiz and on questions of multiple Bloom's levels (Figure 4A and Tables 2 and S1). In addition, self-reported confidence varied depending on student preparation for the quiz and student background (Figure 4B and Tables 2 and S1). Confidence and performance were correlated for students partially using either study method, as they showed both lower paper quiz scores and decreased confidence. On the other hand, the use of method 1 and 2 did increase confidence but did not result in a statistically significant increase in quiz score. Confidence was related to the student's research background, as students with research experience (bench or medical) had greater confidence in their understanding of the paper (Figure 4C and Tables 2 and S1). This was surprising, as research experience did not correlate with increased performance on the paper quiz. Returner students and students with a higher GPA reported an increase in confidence but not to a statistically significant degree (Figure 4, D and E, and Tables 2 and S1).

DISCUSSION

As far as we know, this is the first study to demonstrate a longitudinal increase in the ability to understand and critically analyze a paper as a result of a specific classroom intervention. In addition, we present a comprehensive analysis regarding the factors involved in a student's ability to read and dissect primary literature across multiple biology disciplines. The above conclusions were true not only in aggregate, but similar trends were identified when analyzing the data for each specific course (Figure 1B and Tables 2 and S1). This illustrates that successfully reading papers and learning how to critically analyze them is a discipline- and instructor-independent activity.

Of significance, we demonstrated that learning gains acquired from one course were reflected as an increase in paper quiz performance in later quarters (Figure 2A). The CREATE method is one of the most well-documented techniques to improve student primary literature reading abilities in biology education, with data showing that it is beneficial for students in various courses and at different educational levels (Hoskins *et al.*, 2007; Hoskins, 2008; Gottesman and Hoskins, 2013). Despite this, it is not yet known whether the documented gains are still present after leaving the course in which the CREATE method was introduced.

Surprisingly, students with research experience did not score higher on the paper quiz than their peers without research experience. There have been multiple studies illustrating that students who partake in research report improved critical-thinking abilities and an increased interest in science or enroll in graduate school at higher rates (Lopatto, 2004, 2007; Seymour *et al.*, 2004; Russell *et al.*, 2007; Rios-Velazquez *et al.*, 2011). These studies tend to follow well-structured, highly monitored research opportunities, which unfortunately are not representative of many undergraduate research programs, especially at large universities. A research experience can vary greatly, from a student working on an independent project to a student acting as a laborer without any knowledge about how his or her data contribute to the lab's research question. Our results illustrate that research experience alone may be insufficient to increase student comprehension of primary literature and that it is important for faculty, postdoc, or graduate student mentors to focus not only on training students to perform techniques but also to think about their research experiences in the context of the scientific method and the existing literature. In addition, our data highlight the distinction between clinical and bench research. Students working in a clinical setting often are conducting research related to psychology and patient care rather than the molecular underpinnings of a cell. This research, while valuable, may not produce a greater understanding of primary literature discussed in molecular biology–based labs.

Overall, we observed a positive correlation between performance and confidence (Figure 4A). We attempted to dissect the relationship between performance and confidence, which in past studies has produced varying results (Abouserie, 1995; Morgan and Cleave-Hogg, 2002; Valdez *et al.*, 2006). This may be due to the fact that confidence is a very intricate concept, which is dependent on a number of factors. We created the following model, which ties together the relationship between confidence and performance and the role that outside factors have on confidence, based on our data (Figure 5A). We believe confidence to be a function of a student's prior experience, present work in a given course, self-expectation, and perceptions of course content and difficulty of the assigned task. In our model, "prior experience" is the student's ability to read and analyze a primary literature article before enrolling in the course. "Present work" is the learning gains obtained while studying for the paper quiz, and "self-expectation" is the grade that students believe they will earn on the quiz. Together, prior experience and present work dictate actual understanding and thus should correlate with performance on the paper quiz. The term "perception," in our model, includes the stereotypes associated with various student activities. For example, students may perceive that utilizing the prescribed study methods will increase learning, because the instructor recommended them.

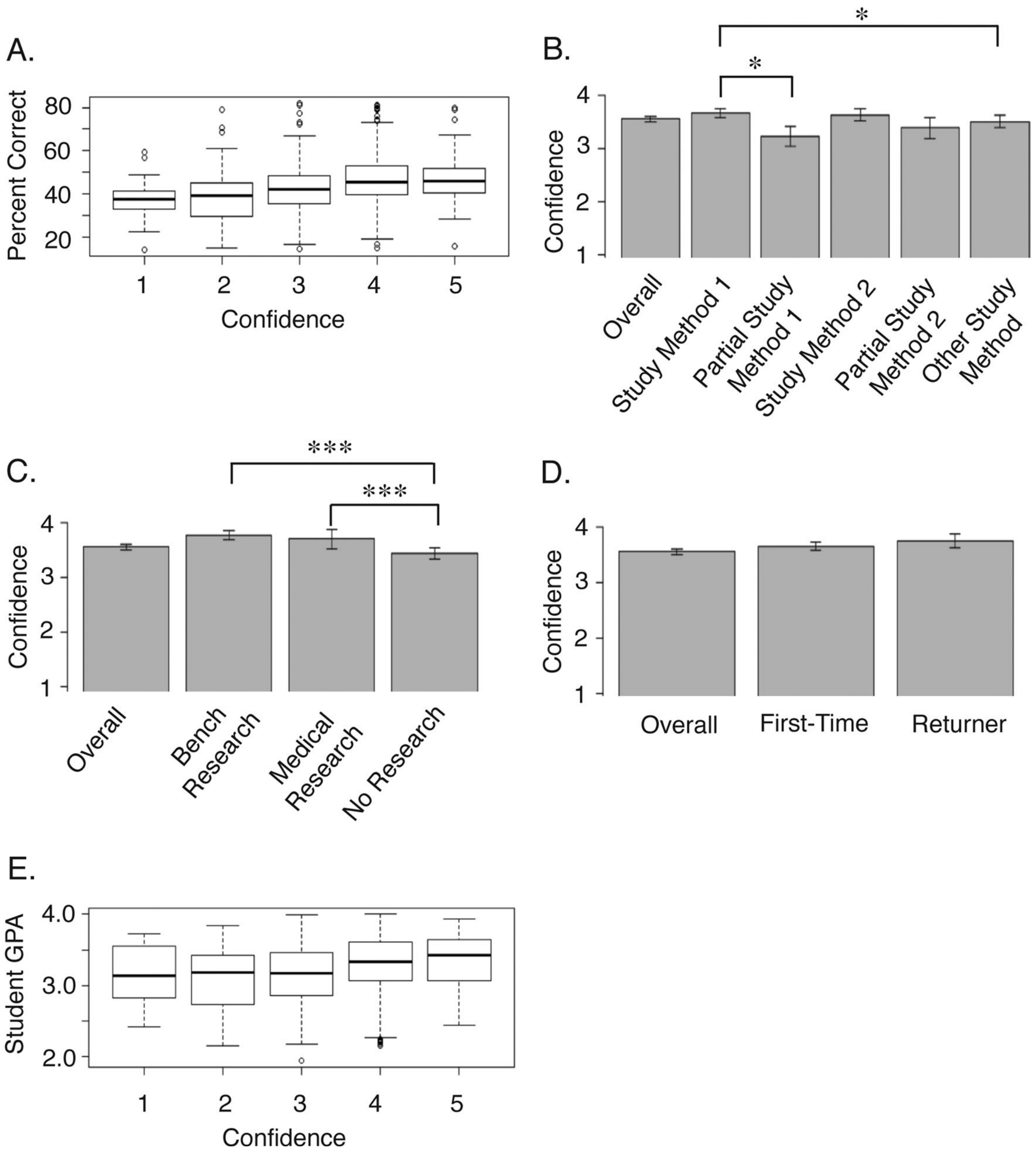

Figure 4. Examining the relationship between student performance on the paper quiz, student background, and self-reported confidence. (A) Overall student scores on the paper quiz are separated based on student self-reported confidence regarding knowledge of the paper. This confidence was reported before taking the paper quiz. Confidence is treated as a continuous variable. Regression analysis shows that confidence is significantly associated with performance ($p < 0.0001$). The top of each box represents the 75th percentile, while the bottom of each box represents the 25th percentile. Whiskers (or inner fences) are positioned at 1.5 times the interquartile range below and above the first and third quartiles. Any points beyond the whiskers are suspected outliers. Postregression diagnostics indicates that suspected outliers are not influential cases and hence do not raise serious concern for the analyses. (B) Self-reported student confidence displayed based on the method of preparation used by the student. (C) Self-reported student confidence displayed based on the student's research experience. (D) Self-reported student confidence displayed based on whether the student is taking one of the included lab courses for the first time. (E) Self-reported student confidence displayed based on the student's GPA. *, $p < 0.05$; ***, $p < 0.001$.

Despite the overall correlation between confidence and performance, this was not evident through all measured parameters. From our study, we observed that students with higher GPAs and returners earned higher scores on the paper quiz yet did not exhibit greater confidence. Students with higher GPAs tend to have increased knowledge (both from previous experience and due to their class preparation) as well as higher expectations, thus these terms will cancel each other out and will produce a similar ratio to low-performing students (Figure 5B). Similar conclusions can be made regarding returner students (Figure 5B). Our data then imply that the perception is that neither group has an advantage; thus the perception value and overall confidence do not increase. On the other hand, study method did impact confidence,

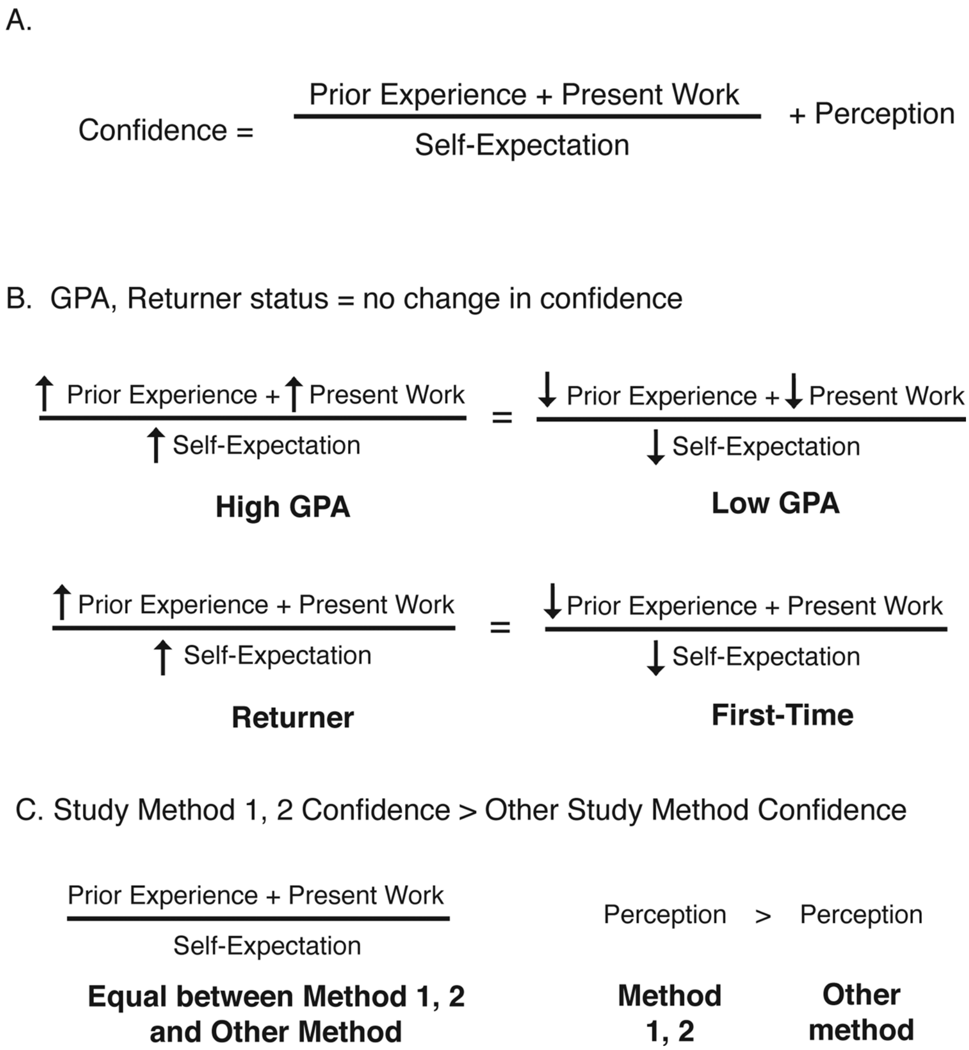

Figure 5. Confidence model. (A) Model describing the relationship between student confidence, performance on a primary literature–based quiz, and educational background. (B) Proposed model to explain why confidence did not change based on GPA and returner status. (C) Proposed model to explain why confidence changed based on study method.

mainly, we believe, due to the perception that instructor-recommended techniques must be beneficial (Figure 5C).

Those with a research background displayed higher confidence, yet there was no significant difference in performance. Students with research experience likely believe that they should be superior at reading scientific literature, increasing the perception value and overall confidence. But students with medical research are unlikely to have covered relevant material, so despite the increase in perception, the prior experience factor for medical research students is lower. This explains why confidence is high with medical research, yet performance is unaffected. Students without research experience, on the other hand, will have a lower perception and thus decreased confidence, which may result in an increased work ethic, a higher present work value, and thus higher performance compared with research students. Confirmation of this model will require follow-up studies in which we will include questions on the paper quiz that are capable of confirming our assumptions regarding prior experience, present work, self-expectation, and perception for students of all backgrounds.

The longitudinal gains we have established stress the importance of focusing on program-wide changes, rather than the single interventions that have been shown to increase learning in the short term. Although more than 90% of our students reported encountering scientific papers in previous courses, these experiences tended to consist primarily of students being assigned papers to complement their in-class or textbook learning with no follow-up assessment or discussion on methods for approaching primary literature. Altering the lab curriculum was beneficial for student literature comprehension and analysis skills, and expanding this to lower- and upper-division lecture courses will likely produce even more positive benefits. By creating program-wide change to augment student exposure to primary literature, we can increase our students' higher-order thinking abilities, which are essential to all STEM fields.

ACKNOWLEDGMENTS

This work was supported by an Assessment Grant from the University of California, Irvine, Assessment, Research & Evaluation Group. We thank A. Nicholas, D. O'Dowd, J. Shaffer, and A. Williams for constructive feedback regarding the manuscript.

REFERENCES

Abouserie R (1995). Self-esteem and achievement motivation as determinants of students' approaches to studying. Stud High Educ *20*, 19–26.

Athanassiou N, McNett JM, Harvey C (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. J Manag Educ *27*, 533–555.

Bissell AN, Lemons PP (2006). A new method for assessing critical thinking in the classroom. BioScience *56*, 66–72.

Clark IE, Romero-Calderón R, Olson JM, Jaworski L, Lopatto D, Banerjee U (2009). "Deconstructing" scientific research: a practical and scalable pedagogical tool to provide evidence-based science instruction. PLoS Biol *7*, e1000264.

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.

DebBurman SK (2002). Learning how scientists work: experiential research projects to promote cell biology learning and scientific process skills. Cell Biol Educ *1*, 154–172.

Fuller D (1997). Critical thinking in undergraduate athletic training education. J Athl Train *32*, 242.

Gore PA (2006). Academic self-efficacy as a predictor of college outcomes: two incremental validity studies. J Car Assess *14*, 92–115.

Gottesman AJ, Hoskins SG (2013). CREATE cornerstone: introduction to scientific thinking, a new course for STEM-interested freshmen, demystifies scientific thinking through analysis of scientific literature. CBE Life Sci Educ *12*, 59–72.

Hoskins SG (2008). Using a paradigm shift to teach neurobiology and the nature of science—a C.R.E.A.T.E.-based approach. J Undergrad Neurosci Educ *6*, A40–A52.

Hoskins SG, Lopatto D, Stevens LM (2011). The C.R.E.A.T.E. approach to primary literature shifts undergraduates' self-assessed ability to read and analyze journal articles, attitudes about science, and epistemological beliefs. CBE Life Sci Educ *10*, 368–378.

Hoskins SG, Stevens LM, Nehm RH (2007). Selective use of the primary literature transforms the classroom into a virtual laboratory. Genetics *176*, 1381–1389.

Kozeracki CA, Carey MF, Colicelli J, Levis-Fitzgerald M (2006). An intensive primary-literature–based teaching program directly benefits undergraduate science majors and facilitates their transition to doctoral programs. Cell Biol Educ *5*, 340–347.

Krathwohl DR (2002). A revision of bloom's taxonomy: an overview. Theory Pract *41*, 212–218.

Livengood J (1992). Students' motivational goals and beliefs about effort and ability as they relate to college academic success. Res High Educ *33*, 247–261.

Lopatto D (2004). Survey of undergraduate research experiences (SURE): first findings. Cell Biol Educ *3*, 270–277.

Lopatto D (2007). Undergraduate research experiences support science career decisions and active learning. CBE Life Sci Educ *6*, 297–306.

Luckie DB, Maleszewski JJ, Loznak SD, Krha M (2004). Infusion of collaborative inquiry throughout a biology curriculum increases student learning: a four-year study of "Teams and Streams." Adv Physiol Educ *28*, 199–209.

Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE Life Sci Educ *9*, 435–440.

Morgan PJ, Cleave-Hogg D (2002). Comparison between medical students' experience, confidence and competence. Med Educ *36*, 534–539.

Rios-Velazquez C, Williamson LL, Cloud-Hansen KA, Allen HK, McMahon MD, Sabree ZL, Donato JJ, Handelsman J (2011). Summer workshop in metagenomics: one week plus eight students equals gigabases of cloned DNA. J Microbiol Biol Educ *12*, 120–126.

Round JE, Campbell AM (2013). Figure facts: encouraging undergraduates to take a data-centered approach to reading primary literature. CBE Life Sci Educ *12*, 39–46.

Russell SH, Hancock MP, McCullough J (2007). Benefits of undergraduate research experiences. Science *316*, 548–549.

Seymour E, Hunter A-B, Laursen SL, DeAntoni T (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. Sci Educ *88*, 493–534.

Stanger-Hall KF (2012). Multiple-choice exams: an obstacle for higher-level thinking in introductory science classes. CBE Life Sci Educ *11*, 294–306.

Valdez CA, Thompson D, Ulrich H, Bi H, Paulsen S (2006). A comparison of pharmacy students' confidence and test performance. Am J Pharm Educ *70*, 76.