

## Article

# The Grass Isn't Always Greener: Perceptions of and Performance on Open-Note Exams

Brian K. Sato,<sup>\*†</sup> Wenliang He,<sup>‡</sup> Mark Warschauer,<sup>‡</sup> and Pavan Kadandale<sup>\*†</sup>

<sup>\*</sup>Department of Molecular Biology and Biochemistry and <sup>‡</sup>School of Education, University of California, Irvine, Irvine, CA 92697

Submitted August 12, 2014; Revised January 26, 2015; Accepted January 27, 2015  
Monitoring Editor: Marshall Sundberg

Undergraduate biology education is often viewed as being focused on memorization rather than development of students' critical-thinking abilities. We speculated that open-note testing would be an easily implemented change that would emphasize higher-order thinking. As open-note testing is not commonly used in the biological sciences and the literature on its effects in biology education is sparse, we performed a comprehensive analysis of this intervention on a primary literature-based exam across three large-enrollment laboratory courses. Although students believed open-note testing would impact exam scores, we found no effect on performance, either overall or on questions of nearly all Bloom's levels. Open-note testing also produced no advantage when examined under a variety of parameters, including research experience, grade point average, course grade, prior exposure to primary literature-focused laboratory courses, or gender. Interestingly, we did observe small differences in open- and closed-note exam performance and perception for students who experienced open-note exams for an entire quarter. This implies that student preparation or in-test behavior can be altered by exposure to open-note testing conditions in a single course and that increased experience may be necessary to truly understand the impact of this intervention.

## INTRODUCTION

Over the past decade, a number of reports have stressed the importance of altering the manner in which science, technology, engineering, and mathematics (STEM) education is conducted at the undergraduate level (National Research Council, 2003, 2009, 2015; Labov *et al.*, 2010; American Association for the Advancement of Science, 2011). Suggested interventions include integration of key concepts throughout the curriculum, increased interdisciplinary course work, development of quantitative skills, and a greater emphasis

on authentic research experiences. These reports were constructed in part due to the widening achievement gap for first-generation, low-income, and underrepresented minority students (Wyner *et al.*, 2007; Chen, 2013; Chang *et al.*, 2014).

Related to these reports, a curricular issue commonly observed in biology classrooms is the emphasis on memorization as opposed to the development of critical-thinking and analytical skills (Seymour and Hewitt, 1997; Wood, 2009; Momsen *et al.*, 2010). While practices to increase critical-thinking skills have been developed (Hoskins *et al.*, 2007; Quitadamo *et al.*, 2008; Sato *et al.*, 2014), these often require very specific methods of implementation or are geared toward certain biology disciplines or classroom environments. For example, both the CREATE method (Hoskins *et al.*, 2007) and Figure Facts (Round and Campbell, 2013) have been implemented only in small-enrollment courses.

To this end, we wanted to examine the effects of open-note testing, an intervention that could be easily introduced into any classroom. Open-note testing has the potential to shift student perception that biology is a memorization-based field (Momsen *et al.*, 2013), as has been demonstrated in other disciplines (Eilertsen and Valdermo, 2000). It has been shown that students taking open-note exams are more likely to prepare for such a test by gathering and critically

CBE Life Sci Educ June 1, 2015 14:ar11

DOI:10.1187/cbe.14-08-0121

<sup>†</sup>These authors contributed equally to this work.

Address correspondence to: Brian K. Sato (bsato@uci.edu) and Pavan Kadandale (pavan.k@uci.edu).

© 2015 B. K. Sato *et al.* CBE—Life Sciences Education © 2015 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>).

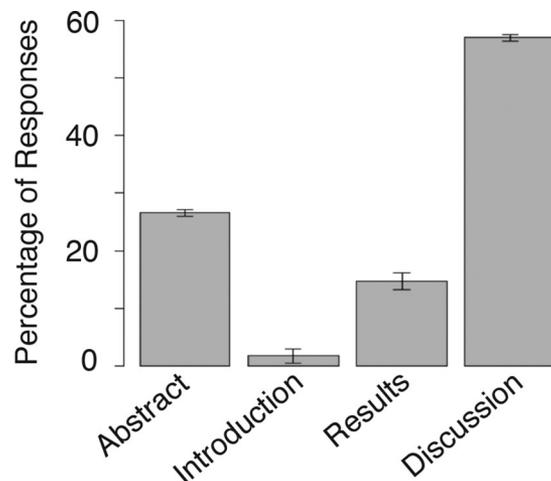
"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

analyzing material from multiple sources (Theophilides and Koutselini, 2000). In addition, students are more inclined to expect that open-note exams will emphasize understanding and analysis (Eilertsen and Valdermo, 2000). On the other hand, students focus on storing information for quick retrieval in preparation for a closed-note exam (Theophilides and Koutselini, 2000). Previous research has produced mixed conclusions in regard to the performance benefits of open-note testing (Koutselini Ioannidou, 1997; Krasne *et al.*, 2006; Agarwal *et al.*, 2008; Agarwal and Roediger, 2011; Heijne-Penninga *et al.*, 2008a,b), although a common finding is that students are more relaxed under these circumstances (Williams and Wong, 2007; Block, 2011; Gharib *et al.*, 2012).

A study regarding the effects of open- versus closed-note testing in a biology classroom examined both student effort and performance in an introductory course (Jensen and Moore, 2008). The authors found that students taking an exam open note outperformed their closed-note peers but, when shifted back to closed-note exams, exhibited decreased performance relative to students who only experienced closed-note testing throughout the course. In addition, students taking open-note exams had lower lecture and review session attendance. One could argue that these results are not particularly surprising, as the tests in this introductory course consisted of fact-based multiple-choice questions, and higher exam performance could have led open-note students to believe less effort was necessary to succeed in the class.

Owing to the lack of consensus regarding open-note testing and the scarcity of studies in biology classrooms, we wanted to explore the effects of this intervention, especially on exams emphasizing critical thinking. The focus of this study was the assessment of scientific paper reading skills in the context of open- and closed-note exams. The ability to understand and critically analyze primary literature is a difficult task that may be compounded by a lack of student metacognition regarding the thought processes required to read a scientific paper. When asked to state the most important section of a paper, the vast majority of our students enrolled in upper-division laboratory courses selected either the abstract or the discussion (Figure 1). Through follow-up conversations, we learned that this potentially was driven by the lack of perceived critical thinking required in biology classrooms. Many students we spoke to viewed scientific articles and textbooks similarly and read both in an analogous manner. This is despite the fact that students typically read a textbook to gather and memorize a set of facts (Guzzetti *et al.*, 1993; Elby, 2001; Randahl, 2012), in contrast to comprehension of primary literature, which requires the higher levels of thinking on Bloom's taxonomy (Krathwohl, 2002; Crowe *et al.*, 2008). Thus, it seemed logical to connect paper reading, a task that involves critical analysis, and open-note testing, which students associate with these higher-order thinking skills (Eilertsen and Valdermo, 2000).

We examined the effects of open- versus closed-note testing on a primary literature-based exam we refer to as the "paper quiz" (see *Methods*). Our study consisted of three main research questions. First, we wanted to determine whether open-note testing provided any performance advantages on an exam requiring both student comprehension of primary literature and critical-thinking abilities. Second, we examined whether these potential advantages were more pronounced under various conditions, including research



**Figure 1.** Student responses to the question, "What is the most important part of a scientific paper?" This question was asked in each of the lab courses during lecture (before the first discussion of a paper), and students answered using the clicker response system.

experience, prior student exposure to our primary literature module, performance in the course, overall grade point average (GPA), and gender. And finally, we examined whether the impact of open-note testing varied based on the degree of exposure to the intervention, by comparing performance on the paper quiz for students who had encountered a series of open-note quizzes over the course of the entire quarter versus those with no prior open-note testing experience.

We answered these questions through an in-depth analysis of the paper quiz guided by Bloom's taxonomy. Questions of different Bloom's levels require different types of thinking, including lower-order processes (Bloom's 1 [recall] and 2 [description]) and higher-order processes (Bloom's 3 [prediction], 4 [analysis], 5 [synthesis], and 6 [evaluation]; Krathwohl, 2002; Crowe *et al.*, 2008). Through Bloom's taxonomy, we assessed student comprehension (Bloom's 1–2 questions) and critical analysis (Bloom's 3–6 questions) skills, as the classifications of thinking represented by the higher Bloom's levels are commonly accepted to be a proxy for measuring aspects of critical thinking (Fuller, 1997; Athanassiou *et al.*, 2003; Bissell and Lemons, 2006; Stanger-Hall, 2012).

Despite student perception that open-note testing would provide a performance advantage, there was no difference in open- and closed-note testing scores overall as well as on questions of most Bloom's levels. Surprisingly, though, when students were given open-note exams over the course of one academic quarter, differences in both perception and performance were uncovered, implying that student test preparation or in-exam behaviors can be altered by exposure to this intervention in a single class.

## METHODS

### *Primary Literature Module Description*

This study was conducted at the University of California, Irvine, a large public research institution. 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 (each course was offered in the Fall, Winter, and Spring quarters). Over the three quarters of the study, enrollments ranged from 40 to 160 students, resulting in paper quiz data from roughly 900 students. Information regarding the data set can be found in Table 1. The biochemistry lab had an approximate enrollment of 40 students per quarter, the molecular biology lab 100 students per quarter, and the microbiology lab 160 students per quarter. Each course consisted of a weekly lecture, which students from all lab sections in that course attended. Students then attended lab sections with an enrollment of 20 per section. The primary literature reading module was only one part of each course, which also required students to take quizzes, write lab reports based on their experiments, and peer review other students' writing samples. Throughout the study period, one instructor taught both the microbiology and biochemistry labs (B.K.S.), while a second instructor taught the molecular biology labs (P.K.). Non-paper quiz exams taken in the microbiology (3 exams) and biochemistry labs (4 exams) were closed note, while those in the molecular biology lab (9 exams) were open note.

In the module, three papers were presented during the quarter. The first two papers were thoroughly discussed in lecture after the students had a week to independently read the paper (Sato *et al.*, 2014). Paper 3 was assigned to the class a week before the paper quiz.

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

### *Instruments (Paper Quiz)*

Data collection for the study occurred through the clicker response system as well as the paper quiz (Sato *et al.*, 2014). Student opinion regarding the most important section of a scientific paper was determined using clickers during course lectures. The paper quiz was a 45-min exam taken in lecture toward the end of the quarter. Paper 3 was presented to students 1 wk in advance of the paper quiz. Paper 3, and hence the paper quiz, was not the same each quarter to minimize any potential advantages students might acquire by referencing previous exams. The paper quiz consisted of questions ranging from Bloom's level 1 to 6 (Crowe *et al.*, 2008). For any figures directly addressed in the paper quiz, the figure and figure legend were provided on the exam (Sato *et al.*, 2014). In addition, each quiz included ungraded questions for students to self-report their independent research background, their confidence in their understanding of the paper as measured on a Likert scale, and their perception of the most advantageous testing method (open vs. closed note). Example paper quizzes along with the Bloom's level for each question have been previously published (Sato *et al.*, 2014). The Bloom's level of each question was determined by a team of three faculty members with experience "Blooming" questions.

In the microbiology and biochemistry labs, the regular, non-paper quiz exams taken throughout the quarter were closed note. In the molecular biology lab, on the other hand, all regular, non-paper quiz exams were administered with an open-note policy. As the paper quiz was toward the end of the course, molecular biology lab students had roughly

7 wk of exams to make potential alterations to their testing practices. We refer to these conditions as "onetime" (microbiology, biochemistry lab) and "chronic" (molecular biology lab) open-note testing.

In each course, half of the students took the paper quiz open note, while the other half took the identical paper quiz closed note. Open-note testing meant that the students could use the research article along with any additional notes. No form of technology (computers, phones, etc.) was allowed for open-note test takers. Students were assigned to the open- or closed-note group by lab section, and each lab section was assigned to a given condition randomly. Students were told whether they would be taking the exam in the open-note or closed-note format a week before the paper quiz (at the same time they were introduced to the paper to read for the quiz). All of the students in a given course took the exam at the same time in the same lecture room, with the room segregated based on testing condition. Students were informed that if paper quiz scores varied based on the intervention, the sections would be graded on separate curves.

### *Statistical Analysis of Data*

Statistical analysis was performed as previously published (Sato *et al.*, 2014). For each of the nine courses included in the study, students were randomly assigned into open- and closed-note testing conditions. The multiple regression technique was used to look for hints of interaction between the experimental condition and demographic variables. While the paper quiz each quarter was distinct, 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 hence were dummy coded on each component level. Specifically, open-note test takers were compared with closed-note test takers; returner students (those who had previously enrolled in one of the study labs) were compared with first-time students (those who were taking one of the study labs for the first time); students with medical research experience or no research experience were compared with students with bench research experience; microbiology lab and biochemistry lab were compared with molecular biology lab; and male students were compared with female students. 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

**Table 1.** Descriptive statistics<sup>a</sup>

	Count	Mean	SD	Minimum	Maximum
<b>Closed note</b>					
Paper quiz score	394	45.00	11.29	14.06	82.02
Bloom's levels 1 and 2	395	79.46	16.22	0.00	100.00
Bloom's level 3	395	53.70	28.56	0.00	100.00
Bloom's level 4	394	17.10	19.10	0.00	100.00
Bloom's level 5	394	14.11	10.96	0.00	50.00
Bloom's level 6	316	32.78	31.99	0.00	100.00
Final grade	395	7.74	2.47	0.00	12.00
Confidence in paper quiz	393	3.65	0.76	1.00	5.00
College GPA	393	3.23	0.42	1.94	3.95
Returner	297	0.18	0.38	0.00	1.00
Female	393	1.47	0.50	1.00	2.00
Onetime	395	0.75	0.44	0.00	1.00
Favor open-note	391	0.61	0.49	0.00	1.00
Favor closed-note	391	0.06	0.24	0.00	1.00
Favor both equally	391	0.33	0.47	0.00	1.00
Bench research	286	0.61	0.49	0.00	1.00
Medical research	286	0.10	0.30	0.00	1.00
No research	286	0.29	0.46	0.00	1.00
<b>Open note</b>					
Paper quiz score	397	45.19	12.41	14.49	86.49
Bloom's levels 1 and 2	398	81.14	14.73	18.18	100.00
Bloom's level 3	398	49.22	28.78	0.00	100.00
Bloom's level 4	398	18.99	22.77	0.00	100.00
Bloom's level 5	398	14.28	11.50	0.00	50.00
Bloom's level 6	317	35.90	34.05	0.00	100.00
Final grade	398	7.80	2.22	0.00	12.00
Confidence in paper quiz	394	3.63	0.72	1.00	5.00
College GPA	396	3.24	0.41	2.18	4.00
Returner	298	0.18	0.38	0.00	1.00
Female	396	1.41	0.49	1.00	2.00
Onetime	398	0.75	0.43	0.00	1.00
Favor open-note	390	0.50	0.50	0.00	1.00
Favor closed-note	390	0.12	0.32	0.00	1.00
Favor both equally	390	0.38	0.49	0.00	1.00
Bench research	291	0.56	0.50	0.00	1.00
Medical research	291	0.12	0.33	0.00	1.00
No research	291	0.32	0.47	0.00	1.00
<b>Total</b>					
Paper quiz score	791	45.10	11.86	14.06	86.49
Bloom's levels 1 and 2	793	80.30	15.51	0.00	100.00
Bloom's level 3	793	51.45	28.74	0.00	100.00
Bloom's level 4	792	18.05	21.03	0.00	100.00
Bloom's level 5	792	14.20	11.23	0.00	50.00
Bloom's level 6	633	34.34	33.05	0.00	100.00
Final grade	793	7.77	2.35	0.00	12.00
Confidence in paper quiz	787	3.64	0.74	1.00	5.00
College GPA	789	3.23	0.41	1.94	4.00
Returner	595	0.18	0.38	0.00	1.00
Female	789	1.44	0.50	1.00	2.00
Onetime	793	0.75	0.43	0.00	1.00
Favor open-note	781	0.55	0.50	0.00	1.00
Favor closed-note	781	0.09	0.29	0.00	1.00
Favor both equally	781	0.36	0.48	0.00	1.00
Bench research	577	0.58	0.49	0.00	1.00
Medical research	577	0.11	0.31	0.00	1.00
No research	577	0.31	0.46	0.00	1.00

<sup>a</sup>Values of Bloom's levels and paper quiz score (a weighted averaged of the Bloom's levels) ranged from 0 to 100%. 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 to 5 for most confident) Likert scale. GPA was on a 4.0 scale. Variables including returner status, gender, chronic exposure to the training, perceived advantages of open versus closed notes, and prior research experience were dummy coded as 0s and 1s. For these variables, the mean value corresponds to the fraction of students found in that category.

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.

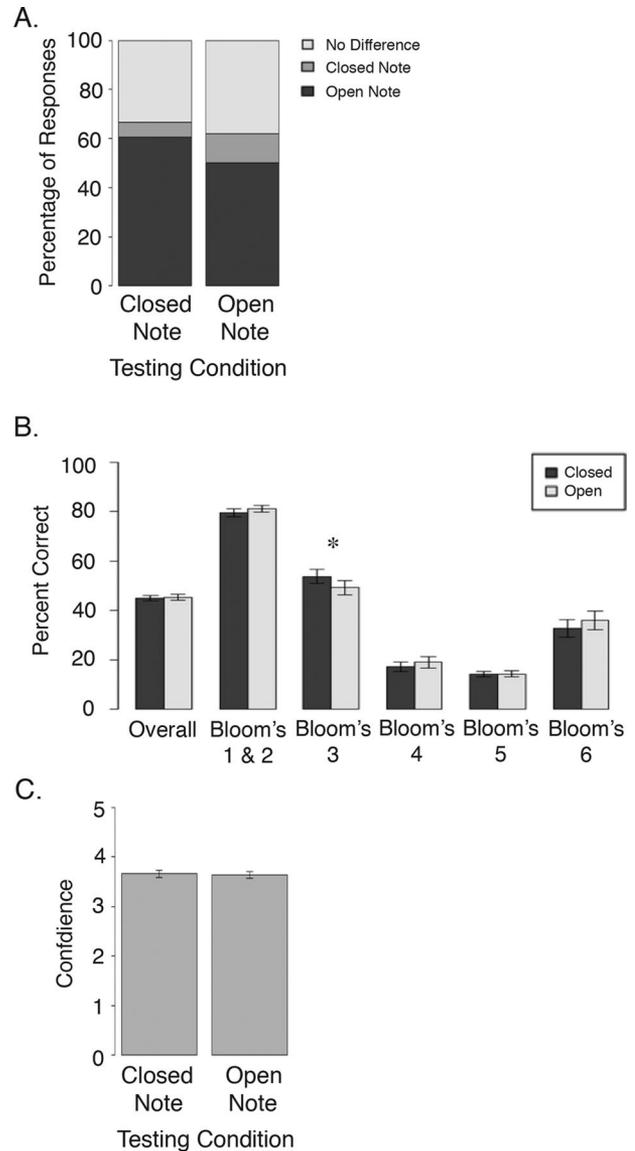
## RESULTS

### *Student Performance in Open- versus Closed-Note Testing Conditions*

Students within a given course were randomly assigned to open- versus closed-note test-taking conditions. No significant differences were found between the groups (open or closed) in regard to overall course grade, gender, GPA, prior exposure to the primary literature module, and research experience (Table 1). Instructors informed the students that the paper quiz would focus heavily on understanding of the scientific article and analysis of figures, rather than memorization of background material or conclusions written by the authors, and that it was not expected that the presence of the article would produce an advantage on the quiz. Despite this, when asked, "What section do you think will have the higher exam mean?," the majority of students selected the open-note takers (Figure 2A). We also observed different answers depending on whether the student was taking the exam open or closed note. The open-note section was more likely to believe that the closed-note students would perform better (11.8% in the open-note section selected this answer vs. 6.1% in the closed-note section,  $p = 0.004$ ). On the other hand, the closed-note section was more likely to believe that open-note testing was an advantage (60.6% in the closed-note section vs. 50.3% in the open-note section,  $p = 0.006$ ).

To determine whether student perception matched reality, we compared overall paper quiz performance as well as performance on questions of specific Bloom's levels between open- and closed-note testing conditions. As seen in Figure 2B, there was no significant difference between the two groups with the exception of questions of Bloom's level 3, for which closed-note students performed better. This was true both of the raw data and when we controlled for observable student characteristics in the regression analysis (Table 2). We speculate that the Bloom's 3 result may be merely noise in the data, both because of the large sample size and the small magnitude of this result (a difference of 0.16 SD). Surprisingly, there was no distinction at Bloom's level 1 and 2, as the presence of the paper might be expected to assist with memorization-based questions.

As the literature has demonstrated that students taking open-note exams are more comfortable (Gharib *et al.*, 2012), we wanted to determine whether this intervention increased student confidence as well. Students were asked to rate their agreement with the following statement "I am very confident I understand the paper being tested" on a five-point Likert scale (5 = strongly agree, 1 = strongly disagree). Despite students' belief that the open-note format conferred an advantage, confidence in understanding, like performance, did not vary whether or not the student was allowed to bring the article to the exam (Figure 2C).



**Figure 2.** Student perception, performance, and confidence in open- versus closed-note testing. (A) Students were randomly split into open- and closed-note testing groups for the paper quiz. Before the paper quiz, students were asked "What section do you think will have the higher exam mean?" Potential answers were open-note section, closed-note section, or that there would be no difference. Answers were recorded on the paper quiz before the students began the exam. (B) Paper quiz scores from nine study lab courses were normalized. Overall paper quiz score and scores on questions of specific Bloom's level were compared between students who took the exam open or closed note. \*,  $p \leq 0.05$ . (C) Before taking the paper quiz, students were asked to report their agreement with the statement "I am very confident I understand the paper being tested" on a five-point Likert scale, with 5 being "strongly agree" and 1 being "strongly disagree." Responses were recorded on the paper quiz and are separated between open- and closed-note test takers.

### *Open- versus Closed-Note Performance Examined under Various Student Classifications*

While there was no overall difference in paper quiz performance between open- and closed-note exam conditions, there may be specific groups of students who were better

**Table 2.** Multiple regression analysis<sup>a</sup>

	(1) Bloom's levels 1 and 2	(2) Bloom's levels 1 and 2	(3) Bloom's level 3	(4) Bloom's level 4	(5) Bloom's level 5	(6) Bloom's level 6	(7) Bloom's level 6	(8) Paper quiz score
Open-note	0.053 (0.138)	0.162* (0.012)	-0.089* (0.016)	0.031 (0.378)	0.049 (0.175)	0.022 (0.464)	-0.083 (0.124)	0.011 (0.740)
Onetime		-0.102 (0.119)					-0.395*** (0.000)	
Open onetime		-0.150* (0.041)					0.145* (0.019)	
Female	-0.047 (0.187)	-0.037 (0.309)	0.030 (0.412)	-0.012 (0.737)	-0.006 (0.867)	-0.063* (0.037)	-0.073* (0.016)	-0.028 (0.387)
Returner	0.088* (0.015)	0.083* (0.023)	0.095* (0.010)	0.054 (0.130)	0.080* (0.028)	0.051 (0.094)	0.056 (0.064)	0.130*** (0.000)
Medical research	-0.071 (0.053)	-0.068 (0.064)	-0.044 (0.241)	-0.081* (0.024)	-0.028 (0.452)	-0.054 (0.079)	-0.057 (0.063)	-0.103** (0.002)
No research	0.003 (0.943)	0.006 (0.885)	0.026 (0.513)	0.013 (0.743)	-0.045 (0.254)	0.012 (0.717)	0.009 (0.780)	0.013 (0.705)
Favor closed	0.014 (0.702)	0.011 (0.775)	0.015 (0.687)	0.013 (0.718)	0.007 (0.849)	-0.017 (0.591)	-0.013 (0.670)	0.016 (0.626)
Favor same	0.012 (0.755)	0.015 (0.681)	-0.019 (0.615)	0.004 (0.910)	0.030 (0.420)	0.014 (0.653)	0.011 (0.735)	0.007 (0.839)
College GPA	0.416*** (0.000)	0.414*** (0.000)	0.266*** (0.000)	0.238*** (0.000)	0.297*** (0.000)	0.155*** (0.000)	0.157*** (0.000)	0.491*** (0.000)
Confidence	0.113** (0.003)	0.112** (0.003)	0.120** (0.002)	0.050 (0.177)	0.042 (0.265)	0.122*** (0.000)	0.123*** (0.000)	0.160*** (0.000)
Observations	571	571	571	571	571	571	571	571
R <sup>2</sup>	0.302	0.308	0.274	0.326	0.300	0.508	0.513	0.443
Adjusted R <sup>2</sup>	0.285	0.289	0.256	0.309	0.283	0.496	0.500	0.429

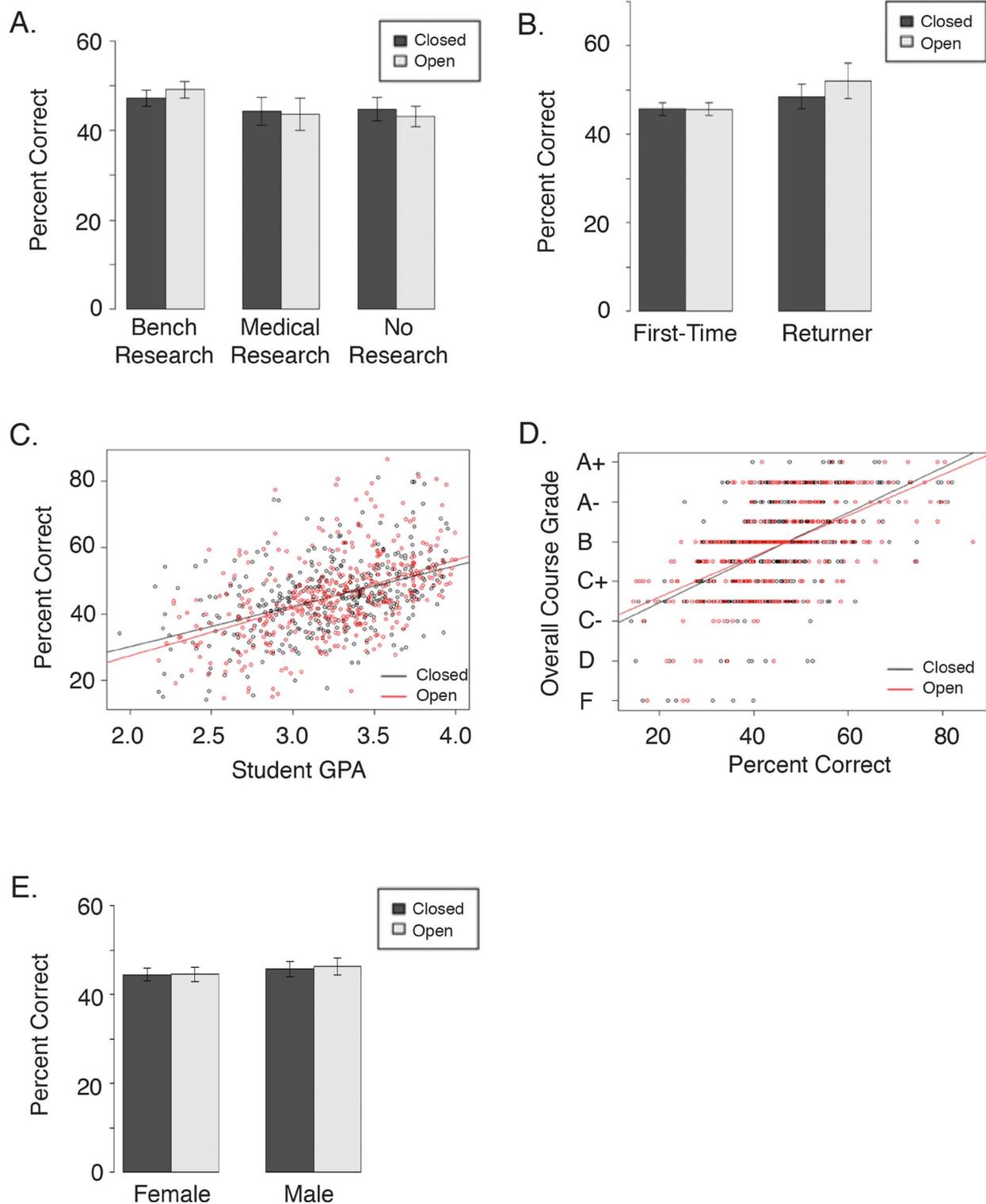
<sup>a</sup>The coefficients in the table are standardized beta coefficients. As a result, they are readily comparable in magnitude across rows and columns, and the coefficients must be interpreted in terms of SDs. Students with bench research experience were used as a baseline for comparing against other students with medical or no prior research experience. Similarly, those who favored the open-note condition served as the comparison group opposite those favoring the closed-note condition or those favoring both equally. 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. Values in parentheses are *p* values. \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$ ; \*\*\*,  $p \leq 0.001$ .

able to take advantage of the scenario compared with their peers. We compared open- versus closed-note paper quiz scores under a variety of parameters, including student research experience, whether the student had enrolled in a prior course with our primary literature module, student GPA, overall course grade, and gender. We have previously demonstrated that returner students (those who had taken a lab course containing our primary literature module) earned higher paper quiz scores than those enrolled in one of the study labs for the first time and that a positive correlation exists between paper quiz performance and overall course grade and student GPA (Sato *et al.*, 2014). Despite these observations, none of the examined factors correlated with a difference in open- versus closed-note testing (Figure 3, A–E, and Table 2).

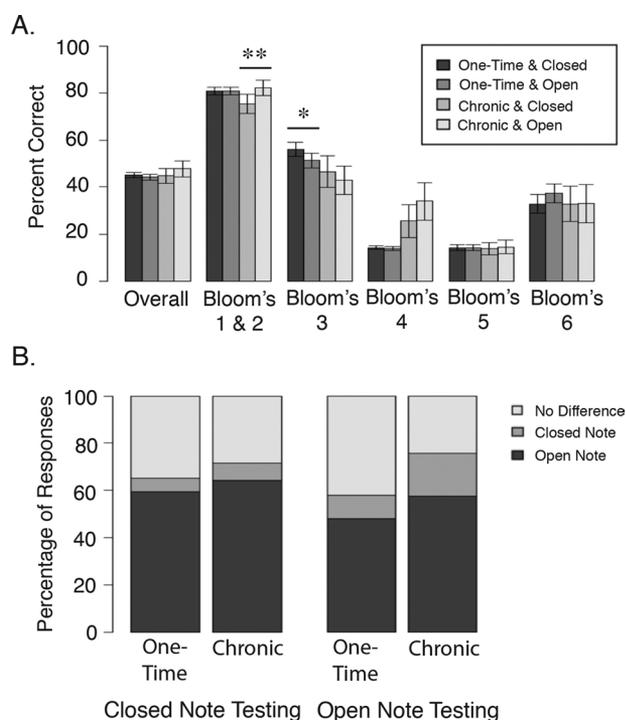
### ***Open- and Closed-Note Performance in Courses with Chronic versus Onetime Open-Note Testing Conditions***

Typically, biology exams are administered under a closed-note environment. Thus, in rare instances in which students are allowed to bring materials with them during the exam, they may be unaware of how to alter their study techniques or in-class behaviors to maximize potential benefits

generated by this aid. To determine whether consistent open-note testing would produce a benefit on the paper quiz, we used two different testing mechanisms (described in *Methods*). Open- and closed-note paper quiz scores were examined under the context of “chronic” exposure to open-note testing (molecular biology lab) versus the “onetime” scenario (microbiology and biochemistry labs). When looking at the overall data (Figure 2B and Table 2), we found no effect of open-note testing, with the exception of students taking the quiz closed note demonstrating higher performance on Bloom's level 3 questions. Examining scores in the context of chronic versus onetime open-note testing uncovered a significant difference in performance on questions at the Bloom's 1 and 2 levels for chronic students and a larger, albeit not statistically significant, difference on overall exam score (Figure 4A and Table 2). While we expected to see the Bloom's 1 and 2 difference for all students, the fact that it is only evident for the chronic students implies that experience with multiple open-note exams is necessary to uncover changes in open-note testing behaviors. In addition, student attitudes shifted with chronic open-note testing (Figure 4B). Chronic open-note testers who were allowed to take the paper quiz open note as well were more likely to believe that closed-note testing was advantageous compared with the views of the onetime group ( $p = 0.02$ ). These results illustrate



**Figure 3.** Overall paper quiz performance comparing open- and closed-note test taking under various parameters. (A) Students self-reported their independent research experience. Paper quiz scores are presented comparing open- and closed-note testing for those with bench research, medical research, or no research experience. (B) Paper quiz scores were compared between students who had previously taken one of the labs involved in this study (returner) versus those taking one of the study labs for the first time (first-time). (C) Paper quiz scores were compared between open- and closed-note test takers based on their GPA. (D) Paper quiz scores were compared between open- and closed-note test takers based on their overall course grade. Letter grades were coded as numbers ranging from 12 ("A+") to 0 ("F"). (E) Paper quiz scores were compared between open- and closed-note test takers based on their gender. No significant difference was observed in open-versus closed-note paper quiz performance under any of the indicated parameters.



**Figure 4.** Differences in paper quiz performance and perception based on chronic versus onetime exposure to open-note test taking. (A) Student data were sorted based on chronic exposure (molecular biology lab) to open-note tests versus those for whom the paper quiz was their only open-note exam (microbiology, biochemistry lab). Overall paper quiz score was then compared between students who took the exam open or closed note. Significance values correspond to the open/closed difference within onetime or chronic testing conditions. \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$  (B) Student responses to the question regarding which section would earn the highest score are noted. Responses are separated based on chronic and onetime open-note exposure.

that modifying the exam structure of just a single course can alter student beliefs and approaches to an open-note exam.

## DISCUSSION

Open-note testing could be a very powerful tool to transform STEM education. Previous work has highlighted the student perception that open-note tests require application and analysis of key concepts, as opposed to the rote memorization for which biology courses are often known (Seymour and Hewitt, 1997; Momsen *et al.*, 2010). Based on a broad survey of undergraduate curricula though, open-note testing is not commonly utilized.

Using a paper-based exam as a means of assessment, we measured student performance on questions of various Bloom's levels in the context of open- and closed-note testing and found little difference when controlling for multiple student parameters (Figure 2B and Table 2). While the literature on the benefits of open- versus closed-note testing is highly variable (Krasne *et al.*, 2006; Heijne-Penninga *et al.*, 2008a), we were particularly surprised about the lack of an impact on Bloom's 1 and 2 questions, which require recall or summarization of

given information. On further reflection, this may be due to the fact that Bloom's 1 and 2 questions can involve facts that are or are not present in one's notes, or in this case, the scientific article. If the answer is not explicitly stated in the text of the paper and the students do not add additional information to their notes while preparing for the exam, the presence of the notes would not be expected to provide an advantage.

On the other hand, chronic exposure to open-note testing uncovered a difference in paper quiz performance on Bloom's 1 and 2 questions when comparing open and closed conditions (Figure 4A and Table 2). It is possible that these experienced students now recognized the limitations of the article text for assistance with fact-based questions and thus constructed more detailed notes during exam preparation. Perception of open- and closed-note testing was also significantly altered between students in the chronic and onetime open-note courses (Figure 4B), further hinting at potential behavioral alterations students make when preparing for or taking an open- or closed-note exam. Based on these results, it appears that an increased frequency of open-note testing in a given curriculum may be necessary to make a firm conclusion regarding this intervention. This may explain the variability in the literature, as different groups of subjects may have encountered open-note testing to different degrees.

The idea that prior experience is important is suggested by a study on an online instructional technology course, in which it was identified that a training module on open-note testing produced exam gains (Rakes, 2008). Added evidence from our work may be seen in the performance of returner students, who by definition have taken at least one more open-note exam than first-time students. Although not statistically significant, the returner group exhibited a greater performance difference between open- and closed-note test takers (Figure 3B). The lack of significance may be due to the fact that the returner group varies in terms of open-note testing experience, as it consists of students previously enrolled in either a chronic or onetime open-note course. If experience is beneficial for open-note testing, we might expect to see differences between chronic returners and onetime returners. Open- versus closed-note performance between these subgroups of returners also leads to insignificant differences (unpublished data), possibly due to the small sample sizes involved. We plan to explore this in future studies.

From our experience, we believe there are a number of practical considerations when thinking about introducing open-note testing into the classroom. As with any technique with which students may be unfamiliar, it is important to explain why it is being introduced. In this case, we stressed that the open-note nature of the test meant success on the paper quiz required independent analysis of the data rather than recitation of the author's text. Further, there must not be a dissonance between the stated objectives of the assessment and the kinds of questions used in the assessment. For example, telling students that the ability to analyze data is important but then designing an exam consisting mostly of lower-level Bloom's questions would send a mixed message. At the very least, we argue that making all exams open note would provide motivation for the instructors to create exams focused on higher-order thinking.

Finally, as this study shows, additional data collection is necessary to determine the benefits, if any, of open-note testing to students. We encourage instructors to introduce

it into their classrooms to help build a consensus regarding this intervention. The appropriate experimental design requires the use of two similar populations of students in a single class taking the identical exam under open- or closed-note conditions. While in our study this was an unpopular decision among students, it is the only way to directly compare potential performance benefits. The most closely related scenario is to compare performance on different exams, one taken open note and the other closed note, with similar content and Bloom's levels. But this introduces potentially problematic variables, as equal Bloom's level is not synonymous with equal difficulty, and student notes may be more beneficial for one set of these similar, yet distinct, questions.

While we did not uncover a significant difference across multiple Bloom's levels in regard to open- and closed-note testing, this is one of the most expansive studies to cover the topic in undergraduate biology education research. Our hope is that an increased number of instructors will consider implementing open-note testing into their classrooms, as the study authors now do for all of their courses. Only through increased exposure to open-note exams and use of well-designed assessments of its impacts will we truly be able to understand the potential benefits of this intervention. In addition, open-note testing forces faculty members to construct exams of higher Bloom's level, providing students with more opportunities to build their analytical skills and helping to erase the stigma that biology is a memorization-based discipline.

## ACKNOWLEDGMENTS

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

## REFERENCES

- Agarwal PK, Karpicke JD, Kang SHK, Roediger HL, McDermott KB (2008). Examining the testing effect with open- and closed-book tests. *Appl Cogn Psychol* 22, 861–876.
- Agarwal PK, Roediger HL (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory* 19, 836–852.
- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC.
- Athanassiou N, McNett JM, Harvey C (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. *J Manage Educ* 27, 533–555.
- Bissell AN, Lemons PP (2006). A new method for assessing critical thinking in the classroom. *BioScience* 56, 66–72.
- Block RM (2011). A discussion of the effect of open-book and closed-book exams on student achievement in an introductory statistics course. *PRIMUS* 22, 228–238.
- Chang MJ, Sharkness J, Hurtado S, Newman CB (2014). What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *J Res Sci Teach* 51, 555–580.
- Chen X (2013). *STEM Attrition: College Students' Paths Into and Out of STEM Fields*, Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- 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.
- Eilertsen TV, Valdermo O (2000). Open-book assessment: a contribution to improved learning? *Stud Educ Eval* 26, 91–103.
- Elby A (2001). Helping physics students learn how to learn. *Am J Phys* 69, S54–S64.
- Fuller D (1997). Critical thinking in undergraduate athletic training education. *J Athletic Train* 32, 242.
- Gharib A, Phillips W, Mathew N (2012). Cheat sheet or open-book? A comparison of the effects of exam types on performance, retention, and anxiety. *Psychol Res* 2, 469–478.
- Guzzetti BJ, Snyder TE, Glass GV, Gamas WS (1993). Promoting conceptual change in science: a comparative meta-analysis of instructional interventions from reading education and science education. *Reading Res Q* 28, 117–159.
- Heijne-Penninga M, Kuks JBM, Hofman WHA, Cohen-Schotanus J (2008a). Influence of open- and closed-book tests on medical students' learning approaches. *Med Educ* 42, 967–974.
- Heijne-Penninga M, Kuks JBM, Schönrock-Adema J, Snijders TAB, Cohen-Schotanus J (2008b). Open-book tests to complement assessment-programmes: analysis of open and closed-book tests. *Adv Health Sci Educ* 13, 263–273.
- Hoskins SG, Stevens LM, Nehm RH (2007). Selective use of the primary literature transforms the classroom into a virtual laboratory. *Genetics* 176, 1381–1389.
- Jensen PA, Moore R (2008). Students' behaviors, grades and perceptions in an introductory biology course. *Am Biol Teach* 70, 483–487.
- Koutselini Ioannidou M (1997). Testing and life-long learning: open-book and closed-book examination in a university course. *Stud Educ Eval* 23, 131–139.
- Krasne S, Wimmers P, Relan A, Drake T (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Adv Health Sci Educ* 11, 155–171.
- Krathwohl DR (2002). A revision of Bloom's taxonomy: an overview. *Theor Pract* 41, 212–218.
- Labov JB, Reid AH, Yamamoto KR (2010). Integrated biology and undergraduate science education: a new biology education for the twenty-first century? *CBE Life Sci Educ* 9, 10–16.
- 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.
- Momsen J, Offerdahl E, Kryjevskaja M, Montplaisir L, Anderson E, Grosz N (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE Life Sci Educ* 12, 239–249.
- National Research Council (NRC) (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academies Press.
- NRC (2009). *A New Biology for the 21st Century*, Washington, DC: National Academies Press.
- NRC (2015). *Reaching Students: What Research Says about Effective Instruction in Undergraduate Science and Engineering* Washington, DC: National Academies Press.
- Quitadamo IJ, Faiola CL, Johnson JE, Kurtz MJ (2008). Community-based inquiry improves critical thinking in general education biology. *CBE Life Sci Educ* 7, 327–337.
- Rakes GC (2008). Open book testing in online learning environments. *J Interactive Online Learn* 7, 1–9.
- Randahl M (2012). First-year engineering students' use of their mathematics textbook—opportunities and constraints. *Math Ed Res J* 24, 239–256.
- 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.

Sato BK, Kadandale P, He W, Murata PMN, Latif Y, Warschauer M (2014). Practice makes pretty good: assessment of primary literature reading abilities across multiple large-enrollment biology laboratory courses. *CBE Life Sci Educ* 13, 677–686.

Seymour E, Hewitt NM (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview.

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

Theophilides C, Koutselini M (2000). Study behavior in the closed-book and the open-book examination: a comparative analysis. *Educ Res Eval* 6, 379–393.

Williams JB, Wong A (2007). Closed book, invigilated exams versus open book, open web exams: An empirical analysis. In: *ICT: Providing Choices for Learners and Learning*. Proceedings Australasian Society for Computers in Learning in Tertiary Education Singapore 2007, ed. RJ Atkinson, C McBeath, SKA Soong, and C Cheers, Centre for Educational Development, Nanyang Technological University, Singapore, 2–5 December, 1079–1083. <http://www.ascilite.org.au/conferences/singapore07/procs/>.

Wood WB (2009). Innovations in teaching undergraduate biology and why we need them. *Annu Rev Cell Dev Biol* 25, 93–112.

Wyner JS, Bridgeland JM, DiIulio JJ Jr (2007). *Achievementtrap: How America Is Failing Millions of High-Achieving Students from Lower-Income Families*, Washington, DC: Civic Enterprises.