

Students' Use of Optional Online Reviews and Its Relationship to Summative Assessment Outcomes in Introductory Biology

Shana K. Carpenter,[†] Shuhebur Rahman,[†] Terry J. S. Lund,[‡] Patrick I. Armstrong,[†] Monica H. Lamm,[§] Robert D. Reason,^{||} and Clark R. Coffman^{†*}

[†]Department of Psychology, [‡]Department of Chemical and Biological Engineering, ^{||}School of Education, and [§]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011; [†]University of Texas Health Science Center, San Antonio, TX 78229

ABSTRACT

Retrieval practice has been shown to produce significant enhancements in student learning of course information, but the extent to which students make use of retrieval to learn information on their own is unclear. In the current study, students in a large introductory biology course were provided with optional online review questions that could be accessed as Test questions (requiring students to answer the questions before receiving feedback) or as Read questions (providing students with the question and correct answer up-front). Students more often chose to access the questions as Test compared with Read, and students who used the Test questions scored significantly higher on subsequent exams compared with students who used Read questions or did not access the questions at all. Following an in-class presentation of superior exam performance following use of the Test questions, student use of Test questions increased significantly for the remainder of the term. These results suggest that practice questions can be an effective tool for enhancing student achievement in biology and that informing students about performance-based outcomes coincides with increased use of retrieval practice.

INTRODUCTION

Successful learning requires students to make a number of decisions regarding what, when, and how to study. These decisions are fundamental to self-regulated learning and can be driven by a variety of factors that are both internal (beliefs and experiences about learning) and external (time and availability of resources). Research from cognitive psychology of learning shows that students' study decisions are not always driven by factors that are diagnostic of actual learning. Rather, they can be influenced by a number of intuitive yet misleading heuristics (e.g., Kornell and Bjork, 2007; Bjork *et al.*, 2013; Finn and Tauber, 2015). Understanding the basis for these decisions is therefore critical to identifying effective study behaviors among students and facilitating students' development of successful self-regulated learning strategies.

A common method of gathering information about students' study behaviors is through survey research. When asked about their study decisions, students often report engaging in behaviors that are not ideal for meaningful long-term learning, such as cramming at the last minute (Hartwig and Dunlosky, 2012). One learning strategy in particular that has been shown to be highly effective but underused by students is self-testing, or retrieval practice. Answering practice questions over to-be-learned information produces significant gains on long-term retention of that information, over and above rereading the information (for recent reviews, see Roediger and Butler, 2011; Carpenter, 2012; Dunlosky *et al.*, 2013; Rowland, 2014; Brame and Biel, 2015). In addition to many demonstrations of the benefits of retrieval in laboratory-based

Michèle Shuster, *Monitoring Editor*

Submitted June 24, 2016; Revised January 31, 2017; Accepted February 3, 2017

CBE Life Sci Educ June 1, 2017 16:ar23

DOI:10.1187/cbe.16-06-0205

*Address correspondence to: Clark R. Coffman (ccoffman@iastate.edu).

© 2017 S. K. Carpenter *et al.* CBE—Life Sciences Education © 2017 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.

settings, more recent research in authentic educational contexts has demonstrated that retrieval practice produces significant gains on middle school students' retention of material in U.S. history (Carpenter *et al.*, 2009) and science (Roediger *et al.*, 2011) and on undergraduate students' retention of information from their courses in psychology (McDaniel *et al.*, 2007, 2012) and biology (Carpenter *et al.*, 2016).

Survey research reveals that retrieval practice does not appear to be widely used by students as a means of learning course material. In one study, Karpicke *et al.* (2009) asked undergraduate students to list the strategies they use while studying and then to rank order the strategies according to how often they used them. The most popular strategy was rereading notes or textbook chapters, with 84% of students listing this strategy, and 55% indicating that it was their most common strategy. In contrast, only 11% of students listed retrieval practice among their strategies, and only 1% listed it as their most common strategy. Other survey studies using a checklist (in which students indicate how often they use each of several listed strategies) have also found that students report rereading their notes more often than they practice recalling the material (Susser and McCabe, 2013).

In addition to survey data on students' self-reported use of study strategies, observational data on students' participation in specific study activities can provide valuable information about the choices they make in a given learning situation. It has sometimes been shown that students' self-reported intentions about their study strategies do not always coincide with their decisions to follow through with those intentions when given the opportunity (Sussan and Son, 2014). Thus, observational data on what students actually do when faced with a decision about how to learn information can be particularly informative.

Observational Data on Students' Use of Retrieval Practice

When given the opportunity, do students engage in retrieval practice as part of their learning of course material? Several studies have offered optional practice quizzes—a form of retrieval practice—before mandatory exams in a course and have found that students do elect to use these practice quizzes. The rate at which they use them, however, is difficult to determine and appears to vary widely across studies and course material. For example, Carrillo-de-la-Peña *et al.* (2009) gave students in six classes at four institutions an optional midterm examination before their mandatory final exam and found that completion rates of the practice midterm ranged from 39% (for first-year psychology students) to 92% (for first-year zoology students). In other studies, the completion rate of an optional practice exam was 46% for first-year dentistry students (Olson and McDonald, 2004), 52% for first-year medical students (Kibble, 2007), and greater than 75% for undergraduate pre-med students (Velan *et al.*, 2008). In a study providing multiple practice quizzes (one true/false and one short-answer quiz over each textbook chapter throughout the semester), Johnson (2006) found that 72% of undergraduate students in an educational psychology course completed at least one quiz, but the average number of quizzes completed was only about 20% of the total number of quizzes available. Johnson concluded that students appeared to be curious about the quizzes and wanted to “check them out” but did not appear to use them regularly to prepare for the mandatory exams.

Johnson's (2006) study suggests that students' participation in optional quizzes is not consistent over time. However, it is unknown when students' participation in the quizzes dropped, as quiz usage in this study was measured as an aggregate number of quizzes accessed across the entire academic term rather than before each exam. Do students show more interest in optional quizzes at the beginning of the term or toward the end? In one recent study, Orr and Foster (2013) found that participation in online quizzes in an undergraduate biology course was highest before the first exam (with nearly 98% of students completing at least one quiz before the exam) but then dropped over the course of the semester (with 85% of students completing at least one quiz before the final exam). The quizzes in this course were required, however, so it is unknown whether the same patterns would be expected for quizzes that are optional.

Understanding students' choices to engage in optional quizzes is important, as participation in these quizzes has been shown to correlate positively with performance on summative assessments. In the studies reviewed earlier that have provided optional quizzes, students who completed the quizzes before the exams scored significantly higher on the exams compared with students who did not complete the quizzes. This advantage ranged from 3.2% in the study by Velan *et al.* (2008) to 12% in the study by Carrillo-de-la-Peña *et al.* (2009). Therefore, identifying when students engage in retrieval practice, and how to encourage it, could be critical to improving students' success in their courses.

Exploring Different Types of Review Opportunities

From these studies, it appears that some students voluntarily engage in retrieval practice through optional quizzes. When they do, they perform better on subsequent exams than students who do not engage in the quizzes. However, it is unknown whether different types of review opportunities lead to different rates of participation and subsequent differences in performance outcomes. It could be that students who choose to review course material via *any* method—practice quizzes, or simply reading the material—score higher on exams than students who choose not to review and that this relationship is in part a reflection of the motivation or conscientiousness of the students who are doing the reviewing versus those who are not. It could also be the case that students who choose not to use the practice quizzes do so because they prefer to review the material through reading and not through quizzing. The fairly high rate of self-reported use of study strategies based on rereading (e.g., Karpicke *et al.*, 2009; Susser and McCabe, 2013) could indicate a popular preference for rereading over one that requires students to retrieve information.

Assuming that reviewing the material is more effective than not reviewing it, students who choose to review through reading may be expected to benefit more than students who do not review the material at all. If some students opt to not use practice quizzes because their preference is to review the information through reading rather than through self-testing, then providing a review option that allows students to simply read the material may result in more overall participation in the reviews, and potentially higher exam scores for those students who would have otherwise not used the reviews at all. Currently, there are no known data available on the proportion of students who choose to review using practice questions

Test Question:

Which of the following will most easily pass through a cell membrane?

- Glucose
- Sodium ions
- CO₂ molecules
- Proteins
- I do not know

Read Question:

Which of the following will most easily pass through a cell membrane?

- Glucose
- Sodium ions
- CO₂ molecules (Correct Answer)
- Proteins
- I do not know

FIGURE 1. Test vs. Read modes of optional review. Students chose to access the questions in Test format, wherein they had to retrieve the answer before the correct answer was revealed, or in Read format, wherein they could read the question and answer together. Students who opted for the Test mode were provided with the correct answers immediately after they had completed all of the questions. Students could access the questions as often as they liked and could engage in either type of review.

designed for quizzing versus reading and how those choices coincide with performance on subsequent exams.

The current study was designed to explore students' use of retrieval-based versus reading-based review questions and how the types of questions students choose to engage with (or choose *not* to engage with) coincide with their later exam scores. In a large introductory biology course, students were provided with review questions posted online after each class that pertained to the topic(s) covered in that day's lessons. Students had the option of accessing the questions via Test mode, in which they attempted to answer the questions on their own before receiving feedback of the correct answers, or Read mode, in which they saw the same questions but with the correct answers already supplied (Figure 1). Students were provided the online practice questions after each individual class meeting throughout the semester. Completing the questions was optional, no points were awarded for completion, and students could access either form of questions (Test vs. Read) as often as they liked.

Consistent with previous research on the use of practice quizzes, we expected that students who accessed practice questions as Test questions would perform higher on subsequent exams compared with students who did not access practice questions at all (Olson and McDonald, 2004; Johnson, 2006; Velan *et al.*, 2008; Carrillo-de-la-Peña *et al.*, 2009). If this advantage is driven primarily by review of the material—regardless of the format of the review (Test vs. Read)—then a similar advantage might be expected for students who access practice questions as Read only. On the other hand, if retrieval practice confers benefits over and above Read-only questions, then accessing practice questions as Test questions, rather than as Read-only questions, would be more likely to coincide with higher exam scores.

Encouraging Participation in Online Reviews

An additional question explored in this research was whether or not student participation in the review questions could be increased through incentives. Previous research has explored the effects of incentives for increasing participation in online quizzes. In a study by Kibble (2007), medical students in a physiology course completed online quizzes that were either completely optional or were graded such that each of four quizzes was worth up to 2% of the final course grade. Only 52% of the

students completed the quizzes when they were optional, compared with nearly 100% when they were graded. Importantly, however, the advantage of improved exam scores for students who completed quizzes versus those who did not complete quizzes was present only under conditions in which the quizzes were optional (80.2% vs. 75.7%, respectively). When the quizzes were graded and each was worth 2% of the final course grade, a much higher number of students completed them, but the subsequent exam scores for students who completed the quizzes did not differ from those of the students who did not complete them (76.1% vs. 77.1%, respectively). Kibble also noted that, when quizzes were incentivized by awarding points for performance, many

students performed well on the quizzes on the first attempt but not necessarily on subsequent exams, raising the possibility that students completed the quizzes primarily as a means of earning the points but not necessarily as a tool to learn the material.

Exploring a different type of incentive system, Kibble (2011) gave students an optional online quiz before each of six mandatory exams in a two-semester undergraduate human physiology course. No credit was given for the quizzes. Instead, to promote use of the quizzes, students were given an orientation lecture on the nature and purpose of formative assessments. This included showing them previous data on the correlation between quiz participation and exam scores and emphasizing that nonparticipation may be linked with poor exam scores. Students were also given regular in-class reminders about the quizzes, along with in-class or online discussions about them. Under these conditions, 88% of students participated in the quizzes (compared with only 52% in the previous study that did not include these promotional activities; Kibble, 2007). Students who participated in the quizzes scored 13% higher, on average, on the mandatory exams compared with students who did not participate. Though the reasons behind the effects of these different incentive systems are unknown (in particular, why students who completed graded practice quizzes in the study by Kibble [2007] did not score higher on subsequent exams compared with students who did not complete them), the fairly high rate of participation and superior exam scores for students who completed no-stakes practice quizzes following the promotion of the quizzes in Kibble's (2011) study might suggest that knowledge of the importance of practice quizzes could increase their use and effectiveness as a learning tool.

The current study used a new approach to encourage participation in the online reviews, providing students with knowledge of their own performance-based outcomes. At the beginning of the term, students were introduced to the online review questions but were not encouraged to use one method or the other (Test vs. Read). Students' choices of which method to use, and how often they accessed the questions, was tracked and correlated with their scores on exams 1 and 2. After exam 2, an in-class demonstration by the instructor showed them the average scores on exams 1 and 2 as a function of the review method that students used. Students' participation in the online

review questions before exams 3 and 4 was then tracked as before. Of particular interest was whether, after seeing the data, students more often opted to use the method(s) that coincided with higher exam scores before the remaining exams.

METHODS

Students and Course

The course, Principles of Biology II, is designed for life science majors and includes material related to cellular, molecular, and chemical bases of life, as well as the structure and function of plant, animal, and microbial systems. There were 317 students enrolled, and the course met twice per week for 80 minutes. There were four mandatory exams. This study was reviewed by the Iowa State University Institutional Review Board and was determined to be exempt.

Online Reviews

For each day's class, a set of multiple-choice review questions was created that aligned with learning outcomes for that day. An undergraduate research assistant who had knowledge of the content attended each class and drafted the questions, ensuring that the questions aligned with the lessons each day. The instructors then edited the questions for form and clarity before they were posted online.

Between three and seven review questions were constructed for each day's class, and these were posted online using the course management system. The questions were posted within a few days after each class and remained available to students throughout the rest of the semester. The questions could be accessed via a link on the course website. From that link, students could access all of the review questions organized by day and topic (e.g., "January 15: Macromolecules"). Within each day/topic, students were provided with two different links: 1) Answer questions from January 15 and then get feedback (accompanied by the instructions "Test your knowledge of concepts from January 15. After you answer all of the questions, the correct answers will be provided"), or 2) Read questions with answers from January 15 (accompanied by the instructions "Read questions with answers provided over concepts from January 15, without testing yourself first"). Five reviews were available (corresponding to five class meetings) that covered content assessed on exam 1, six reviews that covered content assessed on exam 2, and seven reviews each that covered content assessed on exams 3 and 4. All questions were in multiple-choice format, with four alternatives plus an option to indicate "I do not know." Each review could be accessed via Test or Read.

An example of a Test versus Read question is shown in Figure 1. For the Test questions, students received feedback on the correct answer after answering all the questions. The feedback screen for each question was identical to the screen for the Read format, in which the correct answer was highlighted in green type. Thus, for both Test and Read questions, students received the same display of the correct answer. The only difference was whether they tried to answer the question first (Test) or were given the answer up-front (Read).¹

¹The Test questions followed by feedback necessarily involved two separate presentations of the question—first alone, and then in the context of the correct answer—whereas the Read questions involved only one presentation. The benefits of testing over reading readily occur in experimental studies that control the

The multiple-choice review questions were introduced to students by the instructor during the first week of class. Instructions were also posted on the course website informing the students that the review questions were optional (no points assigned), the questions would be available throughout the semester, and that students could access them as often as they liked. Before each of four mandatory exams in the course, we tracked how often students accessed the review questions, and in which format they accessed the questions (Test or Read). The frequency and format with which students accessed the questions was then compared with their scores on subsequent exams.

RESULTS

The following analyses addressed four specific questions: 1) What method(s) do students use to complete online reviews? 2) How does the choice of review method coincide with later exam performance? 3) How does the frequency of accessing online reviews coincide with later exam performance? 4) Does knowledge of exam performance coincide with increased participation in optional online reviews? All analyses are based on the 301 students who completed all four mandatory exams (excluding nine students who dropped the course, and seven additional students who did not complete all four exams).

What Method(s) Do Students Use to Complete Online Reviews?

We measured whether students chose to complete the online reviews using the Test method, the Read method, some combination of Test and Read methods, or whether they did not complete any of the online reviews. Our primary interest was students' interactions (or lack thereof) with the review questions as preparation for each of the four exams. Thus, we tracked only the number of times students completed the reviews *before* taking the exam that covered the content in those reviews (i.e., students' use of the five reviews before exam 1, the six reviews before exam 2, and the seven reviews each before exams 3 and 4). We also determined the proportion of students who completed all of the reviews that were available before each exam, and the proportion who completed at least one review before an exam but did not complete all of them.

As shown in Figure 2, approximately half of the class did not complete any of the reviews before exam 1 and exam 2, with 48% and 42% of students, respectively, having not accessed the reviews at all before these two exams. Following the in-class demonstration showing exam performance as a function of review use (to be detailed in the following sections), the proportion of students who did not access the reviews decreased to 27% and 21% before exam 3 and exam 4, respectively. Thus, by exam 4, nearly 80% of the class completed at least one of the reviews before each exam. For students who completed the reviews, the most popular choice of review method was Test, followed by a combination of Test and Read, followed by Read.

Participation in the reviews increased over the course of the semester, but the largest increase occurred between exam 2

number of presentations and amount of exposure to the material (e.g., Roediger and Karpicke, 2006; Roediger and Butler, 2011), suggesting that time on task is not believed to be the mechanism that underlies these benefits.

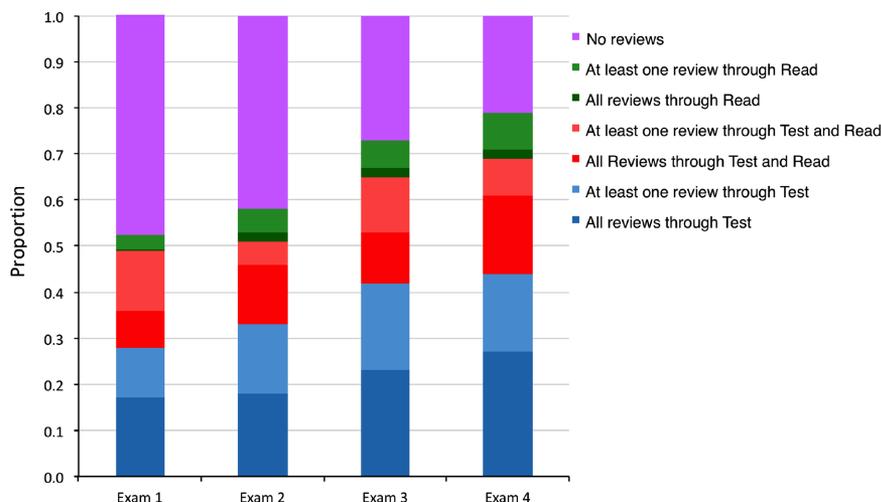


FIGURE 2. Proportion of students completing the reviews through Test, Read, some combination of both, or neither. The proportion of students not using the optional reviews decreased between exams 2 and 3 (No reviews bar, purple, top). Revealing the relationship between review choice and exam performance coincided with a significant increase in the use of Test reviews (blue bars, bottom), but no significant increase in the use of Read reviews (green) or a combination of Read and Test reviews (red).

and exam 3. A McNemar test revealed a significant increase in the proportion of students who used the Test reviews (this includes students who completed anywhere between one review and all reviews) between exam 2 (33% of students) and exam 3 (43% of students), $\chi^2 = 8.95, p = 0.003$. The same analysis revealed no significant difference in the proportion of students using Read reviews between exam 2 (7% of students) and exam 3 (8% of students), $\chi^2 = 0.03, p = 0.86$, and no significant

increase in the use of the online reviews occurred between exam 2 and exam 3, and the significant increase was specific to the use of the Test review.

How Does the Choice of Review Method Coincide with Later Exam Performance?

Figure 3 shows the average exam scores as a function of which review method students chose to use. This includes students who used each review method (Test, Read, or a combination of Test and Read) and completed anywhere between one and all of the reviews using that method. Exam performance of students who did not complete any of the reviews (No Reviews) was also tracked.

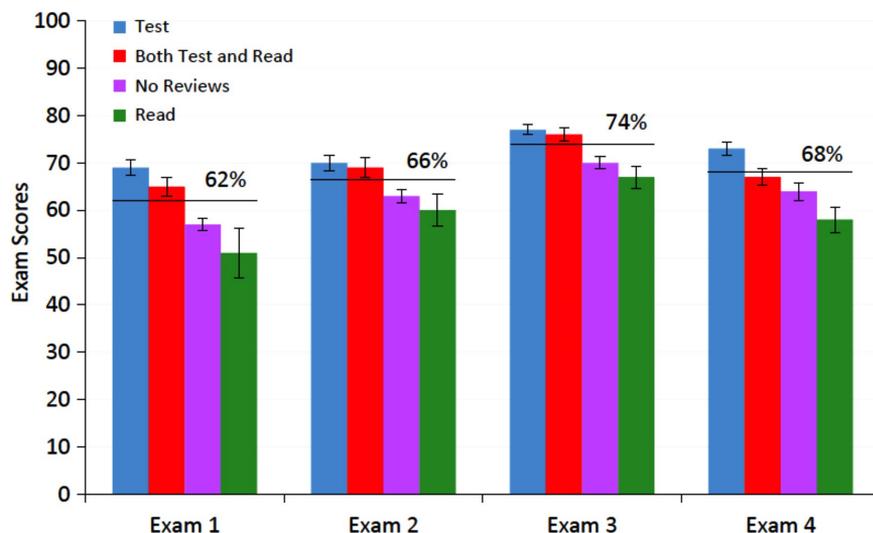


FIGURE 3. Exam performance as a function of review method. Across all four exams, students who used Test reviews (blue, left bar) outperformed students who used Read reviews (green, right bar) and students who did not use the reviews (purple, second from right bar). Students who used Test reviews also scored significantly higher than the class average (depicted by the horizontal line) on each exam. Students using a combination of Test and Read reviews scored higher than those using Read reviews on all four exams. Error bars represent standard errors.

difference in the proportion of students using a combination of Test and Read reviews between exam 2 (18% of students) and exam 3 (23% of students), $\chi^2 = 1.87, p = 0.17$.

Further analysis revealed that the increase in the use of Test reviews was specific to the time interval between exam 2 and exam 3. The proportion of students using the Test review across exams 1 and 2 was quite similar (28% vs. 33%, respectively), $\chi^2 = 1.43, p = 0.23$, as was the proportion using a combination of Test and Read across exams 1 and 2 (21% vs. 18%, respectively), $\chi^2 = 0.44, p = 0.51$.²

The proportion of students using the Test review was also similar across exams 3 and 4 (43% vs. 44%, respectively), $\chi^2 = 0.10, p = 0.75$, as was the proportion of students using the Read review across exams 3 and 4 (8% vs. 10%, respectively), $\chi^2 = 1.16, p = 0.28$, and the proportion of students using a combination of Test and Read across exams 3 and 4 (23% vs. 25%, respectively), $\chi^2 = 0.28, p = 0.59$. Thus, the largest

For the exam 1 scores, a one-way univariate analysis of variance (ANOVA) revealed a significant effect of review method, $F(3, 297) = 13.01, p < 0.001$. Post hoc tests using planned comparisons between the different review methods revealed that students who used the Test review outperformed those who used the Read review, $t(91) = 3.21, p = 0.002$, and those who did not use the reviews, $t(229) = 5.63, p < 0.001$. Students who used a combination of Test and Read also

²Though use of Read reviews increased from exam 1 to exam 2, $\chi^2 = 5.63, p = 0.018$, this result should be interpreted with caution, as the small number of students using Read reviews (only eight students before exam 1, and 22 before exam 2) makes it difficult to obtain a reliable estimate of the magnitude of the difference.

outperformed those who used Read, $t(68) = 2.43$, $p = 0.018$, and those who did not use the reviews, $t(206) = 3.62$, $p < 0.001$.

Scores on the remaining exams largely paralleled these effects. The ANOVA on the exam 2 scores revealed a significant effect of review method, $F(3, 297) = 4.61$, $p = 0.004$, and post hoc comparisons revealed that students who used Test outperformed those who used Read, $t(118) = 2.77$, $p = 0.007$, and those who did not use the reviews, $t(222) = 3.06$, $p = 0.002$. Students who used a combination of Test and Read also outperformed those who used Read, $t(75) = 2.10$, $p = 0.039$, and those who did not use the reviews, $t(179) = 2.01$, $p = 0.045$.

Analysis of exam 3 scores again revealed a significant effect of review method, $F(3, 297) = 10.30$, $p < 0.001$, and post hoc comparisons revealed that students who used Test outperformed those who used Read, $t(150) = 4.06$, $p < 0.001$, and those who did not use the reviews, $t(207) = 4.38$, $p < 0.001$. Students who used a combination of Test and Read also outperformed those who used Read, $t(90) = 3.49$, $p = 0.001$, and those who did not use the reviews, $t(147) = 3.13$, $p = 0.002$.

Consistent with the previous three exams, analysis of exam 4 scores revealed a significant effect of the review method, $F(3, 297) = 10.58$, $p < 0.001$, and post hoc comparisons revealed that students who used Test outperformed those who used Read, $t(161) = 5.03$, $p < 0.001$, and those who did not use the reviews, $t(194) = 3.63$, $p < 0.001$. Students who used a combination of Test and Read also outperformed those who used only Read, $t(103) = 3.11$, $p = 0.002$, but only slightly (and nonsignificantly) outperformed those who did not use the reviews, $t(136) = 1.15$, $p = 0.25$.

Comparing Exam Scores Associated with Each Review Method to the Class Average. Across all four exams, only students who used Test reviews scored significantly higher than

the class average (Figure 3). This was the case for exam 1, $t(84) = 4.012$, $p < 0.001$, exam 2, $t(97) = 2.75$, $p = 0.007$, exam 3, $t(127) = 3.46$, $p = 0.001$, and exam 4, $t(131) = 3.68$, $p < 0.001$. Students who used Read reviews actually scored significantly lower than the class average on exam 1, $t(7) = 3.97$, $p = 0.005$, exam 3, $t(23) = 2.95$, $p = 0.007$, and exam 4, $t(30) = 4.19$, $p < 0.001$. Finally, students who did not use the reviews scored significantly lower than the class average on exam 1, $t(145) = 4.04$, $p < 0.001$, and exam 3, $t(80) = 2.75$, $p = 0.007$.

Exam Scores as a Function of Review Method and Type of Question. On each exam, some of the same questions that had appeared on the online reviews were included, and these were designated as “Repeated Questions.” Other times, some questions from the online reviews were included, with a minor change of wording that rendered the original correct answer incorrect and a different answer correct. For example, the original question, “What type of transport mechanism could be used to move a molecule of glucose across a membrane against a concentration gradient?” was altered by changing the word “against” to “down,” such that the question on the exam read, “What type of transport mechanism could be used to move a molecule of glucose across a membrane down a concentration gradient?” These were designated as “Modified Questions” and appeared only on exams 1 and 3. Finally, each exam also contained a number of questions that had never been seen before by students, and these were designated as “New Questions.”

Table 1 reports the average exam performance for each question type according to the method that students used to complete the online reviews. The scores for each exam were analyzed as a function of which review method students chose (Test, Read, a combination of Test and Read, or No Review) and the type of question that appeared on the exam (Repeated, Modified, or New).

TABLE 1. Exam scores as a function of review method and type of question^a

	Review method				Total
	Test	Read	Test and Read	No Review	
Exam 1					
Repeated questions	81 (17)	57 (26)	78 (20)	60 (22)	70 (23)
Modified questions	57 (39)	19 (26)	46 (38)	39 (37)	45 (38)
New questions	67 (15)	52 (9)	64 (16)	57 (15)	61 (16)
Total	69 (15)	51 (8)	65 (16)	57 (15)	62 (16)
Exam 2					
Repeated questions	91 (17)	81 (24)	91 (16)	75 (24)	83 (22)
New questions	67 (16)	57 (18)	65 (18)	62 (18)	64 (17)
Total	70 (14)	60 (17)	69 (16)	63 (17)	66 (16)
Exam 3					
Repeated questions	87 (14)	77 (13)	87 (14)	70 (16)	82 (16)
Modified questions	90 (13)	76 (18)	87 (16)	80 (19)	86 (17)
New questions	69 (13)	60 (12)	67 (12)	68 (14)	68 (14)
Total	77 (11)	67 (11)	76 (10)	70 (13)	74 (12)
Exam 4					
Repeated questions	86 (16)	74 (18)	82 (17)	64 (18)	79 (19)
New questions	69 (17)	53 (14)	63 (16)	64 (18)	65 (17)
Total	73 (15)	58 (13)	67 (14)	64 (16)	68 (16)

^aScores are expressed as percentages. Standard deviations are given in parentheses.

Analyses of review method for each question type revealed results that were highly consistent with the analysis of overall exam scores. Across all exams and all question types, significant effects of review method were observed, all $F_s > 3.15$, $ps < 0.03$. In all cases, students who used the Test review significantly outperformed students who used Read, $ts > 2.19$, $ps < 0.031$. Students who used Test also generally outperformed students who did not use the reviews, $ts > 2.30$, $ps < 0.025$ (with exceptions for new questions on exam 4, for which the difference was marginal, $p = 0.06$, and new questions on exam 3, for which the difference was not significant, $p = 0.49$). Finally, students who used a combination of Test and Read generally outperformed students who used Read, $ts > 2.20$, $ps < 0.031$ (with marginal effects for modified and new questions on exam 1, $ps = 0.05$, and new questions on exam 2, $p = 0.06$), and students who did not use the reviews, $ts > 2.15$, $ps < 0.035$ (with no significant differences for the modified questions on exam 1, $p = 0.24$, or for the new questions on exams 2, 3, and 4, $ps > 0.22$).

How Does the Frequency of Accessing Online Reviews Coincide with Later Exam Performance?

The total number of times each student completed the online review questions via Test or Read was tracked, and this number was correlated with their scores on the exams following those reviews. These values are reported in Table 2.

The number of Test reviews completed was positively and significantly correlated with exam performance. This effect occurred for the repeated questions across all four exams and

TABLE 2. Correlations between number of reviews completed and exam scores^a

	Test	Read
Exam 1		
Repeated questions	0.47**	-0.02
Modified questions	0.15	-0.07
New questions	0.07	-0.10
Total	0.15	-0.08
Exam 2		
Repeated questions	0.28**	0.18
New questions	0.08	-0.26*
Total	0.11	-0.23*
Exam 3		
Repeated questions	0.54**	0.20
Modified questions	0.20*	-0.10
New questions	0.03	-0.16
Total	0.23**	-0.01
Exam 4		
Repeated questions	0.40**	0.21*
New questions	0.22*	-0.20*
Total	0.29**	-0.10

^aValues are based on students who completed at least one review via Test ($n = 147$) or Read ($n = 70$) before exam 1, at least one review via Test ($n = 153$) or Read ($n = 77$) between exam 1 and exam 2, at least one review via Test ($n = 196$) or Read ($n = 92$) between exam 2 and exam 3, and at least one review via Test ($n = 206$) or Read ($n = 105$) between exam 3 and exam 4. Correlations are based on Spearman's rank-order coefficient, as the number of reviews completed yielded a nonnormal distribution. **Significant at the 0.01 level; *significant at the 0.05 level.

was also present on the modified questions on exam 3 and new questions on exam 4. The number of Read reviews completed, on the other hand, was often *negatively* correlated with exam performance, particularly for new questions. Thus, students who completed more Test reviews tended to score higher on exams, whereas students who completed more Read reviews tended to score lower.

Table 3 shows the average proportion of reviews completed by students before each exam. Students who used Read reviews tended to complete fewer reviews than were available before each exam. For example, the eight students who used Read reviews before exam 1 completed, on average, only about half of the number of reviews available (a proportion of 0.48). On the other hand, students who used Test, or some combination of both Test and Read, tended to complete at least one review more than once. Thus, students who used Read reviews, compared with students who used Test reviews or a combination of Test and Read, tended to complete fewer reviews overall.³

Does Knowledge of Exam Performance Coincide with Increased Participation in Optional Online Reviews?

Approximately 1 week before exam 3, students were given an in-class demonstration by the instructor that consisted of presenting the average (anonymous) exam scores associated with the review methods that students chose (i.e., the graphs in Figure 2 pertaining to scores on exams 1 and 2). During this demonstration, students were reminded about the online review questions, were shown the average exam scores that coincided with the different review methods chosen by students, and informed that students who used the Test reviews scored significantly higher than the class average on both exam 1 and exam 2. No additional discussion of these results followed and class continued.

Following this demonstration, students' use of the Test reviews increased significantly, whereas use of the Read reviews and use of a combination of Test and Read reviews did not significantly change (see Figure 1 and corresponding analyses). Interestingly, use of the Test reviews did not continue to increase but remained relatively stable from exam 3 to exam 4. It appears, therefore, that the strongest increase in the use of Test reviews coincided with the instructor's in-class demonstration showing the advantage in exam scores associated with Test reviews.

DISCUSSION

The current study provides new data on students' choices of online review methods, and how those choices coincide with later exam performance. When given the option to use review

³The higher exam scores associated with Test reviews compared with Read reviews could be partly influenced by the fact that students who used Read reviews completed fewer overall reviews than students who used Test reviews. Even when controlling for the number of reviews completed by restricting the analysis to those students who completed each available review exactly once, however, average exam scores were still higher for students who used Test reviews compared with those who used Read reviews before exam 1 (70% vs. 51%, respectively), exam 2 (71% vs. 47%, respectively), exam 3 (82% vs. 53%, respectively), and exam 4 (78% vs. 53%, respectively). Though small sample sizes preclude definitive statistical comparisons (the number of students completing all of the reviews through Read ranged from just one student before exam 1 to five students before exam 4), this pattern is consistent with the overall analysis showing that students who use Test reviews score higher on exams than students who use Read reviews.

TABLE 3. Frequency with which students completed reviews before each exam^a

Review method	Exam 1		Exam 2		Exam 3		Exam 4	
	<i>n</i>	Prop	<i>n</i>	Prop	<i>n</i>	Prop	<i>n</i>	Prop
Test	85	0.88	98	1.23	128	1.23	132	1.44
Read	8	0.48	22	0.86	24	0.82	31	0.88
Both	62	1.18	55	1.62	68	1.68	74	1.76
Neither	146	—	126	—	81	—	64	—

^a*n* indicates the number of students out of the total sample (301 students) who used a particular review method; “Prop” indicates the average number of reviews completed per student divided by the number of reviews available before each exam (five reviews for exam 1, six for exam 2, and seven each for exams 3 and 4). Values greater than one indicate that students completed at least one review twice.

questions designed as Test (where the answers were not provided until students answered all of the questions—i.e., retrieval practice) versus Read (where the answers were provided simultaneously with the questions), students more often opted to use the Test questions or some combination of both Test and Read compared with Read only.

Furthermore, students who used the Test reviews scored significantly higher on subsequent exams compared with students who used the Read reviews and students who did not use the reviews at all. This finding is consistent with a number of studies showing that participation in formative assessments (i.e., practice quizzes) coincides with higher exam scores (Olson and McDonald, 2004; Johnson, 2006; Kibble, 2007, 2011; Velan *et al.*, 2008; Carrillo-de-la-Peña *et al.*, 2009; Orr and Foster, 2013; Gibson, 2015). However, the current study shows that this benefit is specific to the type of review, consisting of test-based questions, and not simply to reviewing the material by reading the questions and answers together.

This finding is also in line with many studies on the benefits of retrieval practice, showing that learning information through retrieval (compared with simply reading the information) produces significant benefits on long-term learning and transfer of information (e.g., Roediger and Karpicke, 2006; Roediger and Butler, 2011; Butler and Roediger, 2007; Butler, 2010; Carpenter, 2012; Dunlosky *et al.*, 2013; Rowland, 2014; Brame and Biel, 2015). Despite the fairly robust benefits of retrieval—even in course environments using material from the curriculum (e.g., McDaniel *et al.*, 2007, 2013; Carpenter *et al.*, 2009, 2016; Roediger *et al.*, 2011)—survey research on students’ self-reported study habits has revealed that students seldom use retrieval as a study tool (Karpicke *et al.*, 2009; Susser and McCabe, 2013). Other survey research has shown that, when students do use retrieval-based study techniques (e.g., flash cards, practice questions), the majority of students report doing so in order to check their understanding rather than to directly improve their learning (Kornell and Bjork, 2007; Hartwig and Dunlosky, 2012; Yan *et al.*, 2014). Students’ use of retrieval as a metacognitive tool is encouraging, in that this seems to be a motivating factor for engaging in practice questions. However, using practice questions solely as a check of one’s knowledge would seem to restrict the use of these questions to a limited number of times that students wish to gauge their knowledge. Having direct awareness of the benefits of retrieval would more likely encourage students to engage in repeated retrieval more often, which has been shown to coincide with enhanced benefits of retrieval both in previous research (Carpenter *et al.*, 2008) and in the current study (Table 3). Thus, educating students about the direct benefits of retrieval for enhancing

learning, over and above its use as a metacognitive monitoring tool, is important for helping students improve their self-regulated learning skills.

Given the popularity of rereading as a study strategy reported in survey research (Karpicke *et al.*, 2009; Susser and McCabe, 2013), a reasonable prediction is that students might prefer to engage in review questions for which the answers are provided rather than to answer Test questions. To the contrary, we found that a small portion (less than 10%) of students opted to use the Read-only method. These findings are encouraging in showing that students, when left to their own devices, are more likely to choose the review method that has been shown through empirical research to be more effective. One reason for this apparent discrepancy could be that students in the current study did not necessarily use the review questions as a means of studying the information but rather as a check of their knowledge of the material after they had done their studying. Indeed, survey research showing that students primarily use retrieval as a means of checking their knowledge would seem to support this (Kornell and Bjork 2007; Hartwig and Dunlosky, 2012; Yan *et al.*, 2014). Another reason why students’ self-reports favor rereading could be that opportunities for rereading are ever-present, whereas opportunities for self-testing are not as readily available (i.e., practice questions must either be provided by the instructor or some other resource, or constructed by the students themselves). When practice questions are provided to students in the form of easily accessible, no-stakes review questions, the current study shows that students use these questions more often than simply reading the questions and answers.

Consistent with previous research (Johnson, 2006), we found that students did not make full use of the reviews throughout the semester. Whereas by exam 4 a fairly large percentage of students completed at least one review (79% of students), a smaller percentage (46% of students) completed all of the reviews. Thus, as in Johnson’s study (see also Orr and Foster, 2013), it appears that students were more likely to try some of the reviews rather than to make regular and consistent use of all of them. Nonetheless, it seems quite encouraging that by exam 4 nearly half the class had completed 100% of the reviews. Whether or not this finding is typical for optional online reviews awaits further research, and these results offer a promising starting point for further explorations into students’ use of multiple optional reviews before each exam.

Also encouraging is the finding that students’ use of Test reviews increased significantly following an in-class demonstration showing them the exam scores associated with using Test reviews. Even though students received no credit for completing

the reviews, participation in the Test reviews (but not the Read reviews) increased significantly after seeing these results. Though lack of a control group precludes any definitive cause-and-effect conclusions, this selective increase in participation directly after the in-class demonstration suggests that knowledge of performance-based outcomes associated with a given method might have encouraged use of that method. Previous research has reported similar findings (Kibble, 2007, 2011), and has suggested that such an approach might be more effective for learning than incentivizing participation through assigning course credit.

As in all studies tracking students' voluntary participation in any activity, strong conclusions about the direct effects of Test reviews must be tempered by the fact that students themselves chose which review method to use. The superior exam scores associated with Test reviews could reflect the benefits of retrieval practice, or the characteristics of the students themselves who chose to use this particular method. Also, in the current study, we did not collect data on the reasons why students chose to use a particular review method, and why some students chose to not use the reviews at all. These are interesting questions to be addressed in future research. These caveats notwithstanding, the enhanced achievement associated with Test reviews suggests that online quizzes, for the students who make use of them, can lead to significant learning benefits.

In conclusion, practice questions in the form of easily accessible, no-stakes online reviews are used by students in introductory biology, who seem to prefer the Test method more often than the Read method. The enhanced success on exams following use of the Test method suggests that answering practice questions, over and above simply reading the information, may be uniquely beneficial—perhaps due to both direct (retrieval practice) and indirect (enhanced awareness of one's own knowledge) factors. Overall, these results encourage the use of online practice test questions as effective tools for enhancing student achievement in biology.

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation under grant DUE-1504480 and an Iowa State University Miller Faculty Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or Iowa State University.

REFERENCES

- Bjork RA, Dunlosky J, Kornell N (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annu Rev Psychol* 64, 417–444.
- Brame CJ, Biel R (2015). Test-enhanced learning: the potential for testing to promote greater learning in undergraduate science courses. *CBE Life Sci Educ* 14, es4.
- Butler AC (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cognit* 36, 1118–1133.
- Butler AC, Roediger HL III (2007). Testing improves long-term retention in a simulated classroom setting. *Eur J Cognit Psychol* 19, 514–527.
- Carpenter SK (2012). Testing enhances the transfer of learning. *Curr Dir Psychol Sci* 21, 279–283.
- Carpenter SK, Lund T JS, Coffman CR, Armstrong PI, Lamm MH, Reason RD (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educ Psychol Rev* 28, 353–375.
- Carpenter SK, Pashler H, Cepeda NJ (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Appl Cogn Psychol* 23, 760–771.
- Carpenter SK, Pashler H, Wixted JT, Vul E (2008). The effects of tests on learning and forgetting. *Mem Cognit* 36, 438–448.
- Carrillo-de-la-Peña MT, Baillès E, Caseras X, Martínez À, Ortet G, Pérez J (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Adv Health Sci Educ* 14, 61–67.
- Dunlosky J, Rawson KA, Marsh EJ, Nathan MJ, Willingham DT (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychol Sci Public Interest* 14, 4–58.
- Finn B, Tauber SK (2015). When confidence is not a signal of knowing: how students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educ Psychol Rev* 27, 567–586.
- Gibson SI (2015). Promoting an active form of learning out-of-class via answering online "study questions" leads to higher than expected exam scores in general biology. *PeerJ* 3, e1322.
- Hartwig MK, Dunlosky J (2012). Study strategies of college students: are self-testing and scheduling related to achievement? *Psychon Bull Rev* 19, 126–134.
- Johnson GM (2006). Optional online quizzes: college student use and relationship to achievement. *Can J Learn Technol* 32.
- Karpicke JD, Butler AC, Roediger HL III (2009). Metacognitive strategies in student learning: do students practice retrieval when they study on their own? *Memory* 17, 471–479.
- Kibble J (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Adv Physiol Educ* 31, 253–260.
- Kibble J (2011). Voluntary participation in online formative quizzes is a sensitive predictor of student success. *Adv Physiol Educ* 35, 95–96.
- Kornell N, Bjork RA (2007). The promise and perils of self-regulated study. *Psychon Bull and Rev* 14, 219–224.
- McDaniel MA, Anderson JL, Derbish MH, Morrisette N (2007). Testing the testing effect in the classroom. *Eur J Cogn Psychol* 19, 494–513.
- McDaniel MA, Thomas RC, Agarwal PK, McDermott KB, Roediger HL III (2013). Quizzing in middle-school science: successful transfer performance on classroom exams. *Appl Cogn Psychol* 27, 360–372.
- McDaniel MA, Wildman KM, Anderson JL (2012). Using quizzes to enhance summative assessment performance in a Web-based class: an experimental study. *J Appl Res Mem Cognit* 1, 18–26.
- Olson BL, McDonald JL (2004). Influence of online formative assessment upon student learning in biomedical science courses. *J Dent Educ* 68, 656–659.
- Orr R, Foster S (2013). Increasing student success using online quizzing in introductory (majors) biology. *CBE Life Sci Educ* 12, 509–514.
- Roediger HL III, Agarwal PK, McDaniel MA, McDermott KB (2011). Test-enhanced learning in the classroom: long-term improvements from quizzing. *J Exp Psychol Appl* 17, 382–395.
- Roediger HL III, Butler AC (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn Sci* 15, 20–27.
- Roediger HL III, Karpicke JD (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 17, 249–255.
- Rowland CA (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull* 140, 1432–1463.
- Sussan D, Son LK (2014). Breakdown in the metacognitive chain: good intentions aren't enough in high school. *J Appl Res Mem Cognit* 3, 230–238.
- Susser JA, McCabe J (2013). From the lab to the dorm room: metacognitive awareness and use of spaced study. *Instr Sci* 41, 345–363.
- Velan GM, Jones P, McNeil HP, Kumar RK (2008). Integrated online formative assessments in the biomedical sciences for medical students: benefits for learning. *BMC Med Educ* 8, 52.
- Yan VX, Thai K-P, Bjork RA (2014). Habits and beliefs that guide self-regulated learning: do they vary with mindset? *J Appl Res Mem Cognit* 3, 140–152.