Journal of Applied Research in Memory and Cognition 1 (2012) 18–26

Contents lists available at SciVerse ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study☆

Mark A. McDaniel[a,*], Kathleen M. Wildman[a], Janis L. Anderson[b]

[a] Washington University, St. Louis, USA
[b] Harvard Medical School and Brigham & Women's Hospital, University of New Mexico, USA

ARTICLE INFO

Article history:
Received 10 May 2011
Received in revised form
29 September 2011
Accepted 4 October 2011
Available online 13 October 2011

Keywords:
Testing effect
On-line quizzing
Retrieval practice
Testing effects in a college class

ABSTRACT

The testing effect—a mnemonic benefit of quizzing or testing—has been oft-demonstrated in the laboratory, but rarely evaluated experimentally in the classroom. In a college web-based class, using a within-subjects design, core target concepts were quizzed with multiple choice and short-answer questions, presented for reading, or not presented (no additional-exposure control). Multiple attempts on each quiz (and in the read-control) were encouraged and feedback was available after each. When exam questions were identical to earlier-seen quizzes, short answer and multiple choice quizzes enhanced exam performance over rereading targeted material. When examination questions were different but related to the quizzes, short answer and multiple-choice quizzes conferred benefits over unquizzed target content, and produced exam performance that was nominally better than rereading target material. These experimental results indicate that unsupervised on-line quizzing in a college course enhances exam performance, thereby representing an important extension of laboratory testing effects into the classroom setting.

The testing effect is improved memory for materials after a test or quiz. In laboratory experiments, these memorial benefits extend across different types of material (for a review, see Roediger & Karpicke, 2006a) and include benefits for educationally relevant expository prose (Glover, 1989; Kang, McDermott, & Roediger, 2007; Roediger & Karpicke, 2006b). The benefits of testing can exceed those for rereading or re-study exposure (Carrier & Pashler, 1992; Kang et al., 2007; Roediger & Karpicke, 2006b). Thus, testing is not a neutral event; testing modifies and improves memory (McDaniel & Masson, 1985).

In education, despite the mnemonic benefits of testing, quizzing and testing have typically been used as evaluative tools to assess student performance. This orientation may be changing, however, with the increasing use of on-line quizzing and its facilitation through tools available at many universities (e.g., Blackboard Learning System™). On-line quizzes minimize instructor and class time and administrative costs, can be assigned as learning exercises for students (e.g., as no- or low-stakes formative assessments), and provide immediate feedback to students. Here we investigate whether on-line quizzes could be useful classroom tools to assist student learning in a college course. (We will refer to an interim test as a *quiz*, to distinguish it from the criterial measurement which will be referred to as a *test* or *exam*.)

There are numerous differences between the laboratory and classroom settings, any of which might alter the pattern of testing effects in an actual classroom. For instance in the laboratory, the amount of exposure to material is controlled and usually limited to a single session in which material is exposed once (e.g., Glover, 1989; Kang et al., 2007) or studied in massed fashion prior to quizzing. By contrast, in the classroom setting, exposure to course content is typically repeated, varied, and spaced. This greater exposure to course content may support levels of learning at the upper end of the scale, rather than the intermediate performance levels engineered in experiments. Further, in nearly all laboratory studies the same questions are repeated across initial and final tests (see e.g., Kang et al., 2007; Roediger & Marsh, 2005). In contrast, classroom instructors may avoid using the same questions in quizzes and tests (see Angus & Watson, 2009, for a similar situation in their on-line quizzing study, and Mayer et al., 2009, with in-class quizzes). Presenting different questions on the quizzes from those presented on the final tests could also limit the generality of the testing effects reported in laboratory experiments (Nungester & Duchastel, 1982).

Given the differences between laboratory paradigms and classroom practices, the profile of testing effects in authentic classroom settings is uncertain. A handful of published studies offer

☆ This research was supported by a grant from the Institute of Educational Sciences (No. R305H060080) and by a Collaborative Activity Grant from the James S. McDonnell Foundation (No. 22002004I).

* Corresponding author at: Department of Psychology, Washington University, CB1125, St. Louis, MO 63130, USA. Tel.: +1 314 935 8030.
E-mail address: mmcdanie@artsci.wustl.edu (M.A. McDaniel).

2211-3681/$ – see front matter © 2011 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.jarmac.2011.10.001

preliminary, though not unequivocal, support that the testing effect will extend to classroom settings and materials. For instance, in several recent studies examining the effects of on-line quizzing on college-course performances, students were given the opportunity to access on-line multiple choice quizzes prior to course examinations. Students who did not take the quizzes (Kibble, 2007, in a medical physiology course) or who took fewer quizzes (Angus & Watson, 2009, in an applied mathematics course) scored significantly lower on examinations than did those who took all of the quizzes. Despite the statistical control of some potentially confounding variables (e.g., prior ability and course effort in Angus & Watson), the correlational nature of both studies precludes strong conclusions regarding the causal link between on-line quizzes and examination scores. In a quasi-experimental design, Daniel and Broida (2004) found significant gains on exam scores for an on-line-quiz class section relative to a no-quiz control section when procedures were in place to preclude "cheating" on the quizzes (see Lyle & Crawford, 2011, for a similar benefit of in-class quizzing in a statistics course). These findings are not altogether conclusive because assignment of students to the different sections was not random.

McDaniel, Anderson, Derbish, and Morrisette (2007) provided an experimental demonstration of a positive on-line quizzing effect in an authentic college-course setting. However, this experiment reflected relatively low fidelity with a typical course context because participation was voluntary and quizzed content excluded the facts that would be tested and used to evaluate students for their course grades. In sum, to the best of our knowledge no published experimental study with random assignment has unambiguously established a benefit of on-line quizzes for enhancing performance on course summative assessments (though see Glass, 2009, and Glass, Brill, & Ingate, 2008, for experiments combining on-line and in-class quizzing).

Accordingly, we conducted two controlled experiments using within-subjects manipulations to examine the effectiveness of on-line quizzing (with feedback) as a learning tool for required course content. Our quizzing model incorporated features that have been present in naturalistic studies (as described above) investigating on-line quizzing effects. Most prominently, quizzes could be taken more than once (students were encouraged to take the quizzes four times each) and restrictions were not placed on students in terms of study aids they could access in taking the unsupervised on-line quizzes (see e.g., Kibble, 2007). We implemented two baseline conditions (within-subjects) against which to compare the outcomes for quizzed items. One was a typical *no-quiz* control. The other baseline provided a more stringent comparison for evaluating the benefits of quizzing. This second control presented students with selected content (that would be on the exam) to study (the *read* control). Therefore, we were able to examine whether additional exposure to target material might prove as effective in supporting learning as is exposure through quizzing.

Another objective was to investigate the benefits of quizzing when the quiz questions were identical to the exam questions (*identical* questions) and when the quiz questions focused on information related but not identical to the exam questions (*related* questions). As noted earlier, the laboratory testing-effect paradigms have primarily presented identical questions at quiz and at test. We examined whether these standard laboratory benefits of quizzing (and feedback) relative to both no-quiz and re-study controls (e.g., Kang et al., 2007) would extend to an authentic classroom context and with on-line open-book quizzes (e.g., see Daniel & Broida, 2004, for a quasi-experiment in a college class in which students' use of notes and texts undermined the testing effect).

Less often in laboratory experiments, quizzing has been reported to enhance performance on final test questions that focus on information related to the quiz questions but not explicitly targeted in the quiz questions (Chan, McDermott, & Roediger, 2006). In a classroom experiment, however, Mayer et al. (2009) reported no gains in exam performance after in-class quizzes (without discussion) that had related question to those appearing on the exam relative to a class with no quizzes. But the quizzes were presented only once. Based on quasi-experimental studies that reported benefits of repeating quiz items that were non-identical to exam items (Angus & Watson, 2009; Kibble, 2007), we thought it possible that with repeated quizzing and feedback, the related-quiz condition would produce gains in exam performance relative to the no-quiz control.

Table 1
Counterbalancing of quiz type (exposure) across fact sets and weeks.

Counterbalancing group	Fact Set A			Fact Set B		
	Week 1	Week 2	Week 3	Week 1	Week 2	Week 3
1	MC	SA	Read	NQ	NQ	NQ
2	NQ	NQ	NQ	MC	SA	Read
3	SA	Read	MC	NQ	NQ	NQ
4	NQ	NQ	NQ	SA	Read	MC
5	Read	MC	SA	NQ	NQ	NQ
6	NQ	NQ	NQ	Read	MC	SA

Note. MC: multiple choice; SA: short answer; and NQ: no quiz items.

1. Experiment 1

1.1. Method

1.1.1. Participants

Participants were the students enrolled in a 15-week web-based Brain and Behavior undergraduate college course. All students were required to read textbook chapters weekly (approximately 40 pages each), take weekly on-line quizzes, and take on-line unit tests at the end of each three weeks. As in any classroom, students joined and left the class over the semester, and failed to fulfill some requirements even when enrolled for the entire term. To maximize the number of students contributing data, we analyzed results for all students who took tests in Units 1 and 2 (of five units) and who missed no more than one quiz out of three in either unit. Eighteen students enrolled in the class; data from two students were not included because the students missed an exam and/or missed more than one quiz per unit. Thus, data from 16 participants were used in analyses.

1.1.2. Design

There were two experimental variables, both manipulated within-subjects. One variable, labeled *exposure*, reflected one of four quiz conditions in which a particular target fact could be presented. The four conditions were (1) short answer quiz question, (2) multiple choice quiz question, (3) reading the target fact (*read*; note that this is not a quiz per se but is referred to as such for ease of exposition), and (4) no additional exposure of the target fact. For this experimental variable, two layers of counterbalancing were implemented in assigning each target fact to each of the four exposure conditions across students. First, we did not want the unit exams to be populated with a majority of quizzed (or read) items; otherwise, students might have been biased to only attend to quizzed information. Accordingly, for each chapter (see next section for details), the target facts were divided into two sets with equal numbers of facts in each set (labeled Set A and Set B in Table 1). One fact set served as the no-quiz control set, and the other fact set was assigned to one of the three remaining exposure conditions (short-answer quiz, multiple choice quiz, read). The second layer of counterbalancing was applied to the particular set of facts for each unit (either set A or B) that was assigned to be the exposed set. Each unit covered three chapters (three weeks); for each subject, facts (from the quiz set) from each particular chapter were assigned

to a different exposure condition. The particular facts (chapters) assigned to each quiz condition were counterbalanced across subjects, so that for the experiment as a whole, every chapters' facts appeared in each of the three quiz conditions (short answer, multiple choice, read; see Table 1 for the counterbalancing scheme).

The second experimental variable was the question wording on the unit exams. Half of the unit exam questions repeated the same wording that was used in the quizzes (*identical* questions), and half of the questions targeted information related to that presented on the quizzes. These *related* exam questions tested content that was related to that which had been quizzed but focused on a different aspect of the target concepts (see Appendix A for examples). Whether a particular unit exam item appeared as an identical question (as that on the quiz) or a related question was counterbalanced across students. Note that this question-wording manipulation (on the unit exams) also applied for the items that were seen by students as read items. Thus, the read entry for an identical exam question presented the information needed to answer the question, whereas the read entry for the related question condition presented a companion aspect of the targeted information (but not the answer). For example, for the exam question focusing on efferent flow, the read item, *Information leaving a structure* (*exiting*) *is called efferent*, would represent the identical read condition; by contrast the read item, *Information coming into a structure* (*arriving*) *is called afferent*, would represent the related read condition.

1.1.3. Materials

Twenty-eight facts (concepts) were identified from each chapter of an undergraduate textbook (Kolb & Whishaw, 2006). The facts were chosen by the instructor, and these were the facts that appeared on the course exams. For the read items the entire fact was presented intact (as described above). Multiple choice and short answer quiz questions for each target fact differed only in that a multiple choice question offered alternative answers.

Three chapters (one chapter per week) comprised a unit, and an exam was constructed for each unit. Unit exams consisted of multiple choice questions addressing each of the 84 target concepts from the three-chapter unit (28 target concepts per chapter). The unit exam included 42 questions on material to which students had been exposed at quiz (14 multiple choice questions, 14 short answer questions, and 14 read statements) and 42 questions on material not quizzed (or in the read condition).

1.1.4. Procedure

Students received on-line access to a quiz (recall that for the read condition the "quiz" items were target concepts that students were instructed to read) during the week that the relevant chapter was being covered. They logged onto the course at their discretion to take the 14-question quiz for ungraded credit. Participants had unlimited time to complete each quiz. For the read condition, students responded to each item by clicking, "I have read the above statement." For the short-answer condition students responded by typing the answer. For the multiple-choice condition, students selected one of four possible answers. Access to correct answer feedback became available immediately following the first submission of a quiz. The quiz remained available for retaking until the hour before the unit examination went on-line. Feedback for each quiz condition included the quiz question, answer choices (for the multiple choice quiz), an elaborative statement that contained the answer (and related information that sometimes was directly useful for the related question; see Appendix A for examples of feedback), and the participant's response. The feedback for the read condition displayed the elaborative statement (given for the quiz feedback) that included the information read by the student.

The instructor motivated students to repeat the quizzes four times each by awarding 2.5 points for each quiz submission, up to four submissions per weekly quiz. There was no minimum time that had to elapse before quizzes could be repeated. Points were based on completing the quiz, not on the quiz performance itself. The points for quiz completion across the semester accounted for 15% of the course grade. Students understood that the open-book quizzes were considered an important component of the didactic technique, but were not given any explicit information regarding the relationship between quiz questions and the exam itself, other than the understanding that whereas quiz items were variable in exposure format, the exam items would all be multiple-choice. The unit exams were also on-line but were "closed book" and could only be taken once; a 90-min time limit was imposed for each exam, and students had to take the unit exam in a proctored setting, requiring a password for test access that was given only to proctors.

Table 2
Quiz performance by quiz type for the first four quizzes and the number of students attempting a quiz.

Attempt	Multiple Choice Quizzes			Short Answer Quizzes		
	Mean	SD	*N*	Mean	SD	*N*
Experiment 1						
1st	.50	.20	16	.31	.19	16
2nd	.87	.23	16	.86	.23	16
3rd	.95	.09	15	.94	.10	16
4th	.96	.08	14	.97	.07	16
Experiment 2						
1st	.61	.21	27	.22	.12	27
2nd	.93	.08	27	.83	.21	27
3rd	.97	.05	27	.92	.14	26
4th	.98	.03	27	.93	.14	26

1.2. Results and discussion

We first briefly report the quiz performances and then detail the results of primary interest, the performances on the unit examinations.

1.2.1. Quiz performance

The number of quiz attempts was averaged over the units for each student. A within-subjects analysis of variance (ANOVA) found no significant differences in average quiz attempts across short answer ($M = 3.87$, SD = .99), multiple choice $M = 3.22$, SD = 1.06), and read conditions ($M = 3.38$, SD = .94).

For the 14 students who took a quiz the minimum of four times, performance was averaged by quiz type across units (if a student failed to take a particular quiz type for one unit, then the value was the quiz score from the one unit; see Table 2 for means); one of the authors (KMW) scored the short-answer responses with misspellings allowed. Quiz performances improved from the first attempt to the fourth attempt, $F(3, 39) = 85.17$, $p < .001$, $\eta_p^2 = .87$. A significant interaction between quiz type and quiz attempt, $F(3, 39) = 5.65$, $p < .02$, $\eta_p^2 = .30$, indicated that performance on initial but not later quiz attempts was higher with multiple choice quizzes than for short answer quizzes.

1.2.2. Unit exam performance

Proportions of correct answers across the two unit tests were the dependent variables. Seven scores were computed for each student[1]: (1) Exam performance for the no-quiz control items, (2)

[1] The seven scores per participant were calculated as follows. The no-quiz control score represented performance on 84 criterial exam questions to which a participant had not been exposed in any quiz type (42 questions for each of two units). Scores from exam questions associated with each of the read conditions (identical statements, related statements) were calculated from 14 questions (seven per unit). Similarly, each of the four scores derived from quizzed items

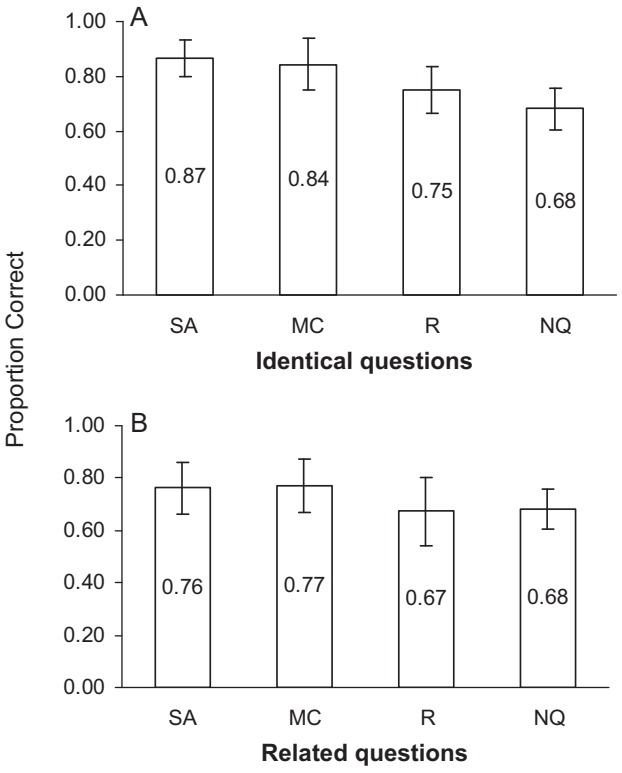

Fig. 1. Mean proportion correct on the unit tests as a function of quiz exposure condition and unit-test question type in Experiment 1 (error bars represent 95% confidence intervals). SA: short answer; MC: multiple choice; R: read; and NQ: no quiz.

performance for the identical read items, (3) performance for the related read items, (4) performance for the short-answer quiz items given the identical exam question and (5) given the related exam question, and (6) performance for the multiple-choice quiz items given the identical exam question and (7) given the related exam question.

We used planned comparisons for the analyses of the unit tests (see Callender & McDaniel, 2007, for a similar approach). We performed four planned comparisons for identical exam question outcomes and four parallel planned comparisons for related exam-question outcomes (the comparisons for the related exam questions were: short-answer related vs. read related and vs. no-exposure control; and multiple-choice related vs. read related and vs. no-exposure control). Finally, we contrasted exam performance on read identical and on read related versus no-quiz items, for a total of 10 comparisons. We used an alpha level of .035 (.35/10) for our planned comparisons to maintain a family wise error rate comparable to that of an omnibus ANOVA approach (which would have included three factors and seven effects). We calculated effect sizes using Cohen's *d*.

Consider first the benefits of quizzing when the stem of the exam questions was *identical* to the quiz questions (see Fig. 1, panel A for means). Both short answer quizzes and multiple choice quizzes enhanced exam performance relative to control (no-quiz) questions, $t(15) = 6.41$, $p < .001$, $d = 1.37$, and $t(15) = 3.28$, $p < .01$, $d = 1.00$, respectively. Similarly, the identical read items enhanced performance over control (no-quiz) items, $t(15) = 2.42$, $p < .029$, $d = .45$. Comparing the short answer and multiple choice quiz conditions (those with identical questions as on the exam) to these read items indicated that only the short answer quizzes significantly enhanced exam performance relative to the read condition; $t(15) = 4.19$, $p < .01$, $d = .83$. Multiple choice quizzes did not produce gains in exam performance that were significantly greater than those produced by reading content that directly answered the exam questions ($t(15) = 2.00$, $p < .07$).

When exam questions were related to the information targeted in the quiz questions (and presented in the read condition), the testing effects remained but reading related content was not helpful (see Fig. 1, Panel B for means). Short answer quizzing and multiple choice quizzing produced significant benefits relative to the no-quiz control, $t(15) = 3.08$, $p < .01$, $d = .50$, and $t(15) = 2.35$, $p < .034$, $d = .52$, respectively. By contrast, reading related information failed to increase performance over no-quiz items, even minimally ($t < 1$). Finally, because the variance associated with these read items was inflated (relative to the no-quiz control items, which included many more observations; see Footnote 1), the benefits of short answer and multiple-choice quizzing relative to the read condition did not reach statistical significance ($t(15) = 1.96$, $p < .07$, and $t(15) = 1.90$, $p < .08$, respectively).

The short-answer quizzing advantage relative to the read items in the identical conditions indicates that the present improvements produced by quizzing cannot be attributed wholly to using print-outs of quizzes as a study guide, or of memorizing the facts highlighted in the quizzes (cf. Daniel & Broida, 2004). The read identical-content items offered adequate opportunities for repetition or for printing a study guide. If students were simply memorizing the information, then the read identical condition should have been equivalent to the short-answer identical quiz condition. This pattern was not observed. Further, the present results establish that unsupervised (allowing access to notes and texts) on-line quizzes with feedback can effectively enhance exam performance, even when the quiz items are not identical to the exam questions. One quasi-experimental study in a classroom suggested otherwise (Daniel & Broida, 2004). The results from the present experiment are based on a modest number of students, however. Accordingly, to reinforce these results, we conducted a second experiment in an eight-week summer course for which enrollment was higher.

2. Experiment 2

2.1. Method

2.1.1. Participants and design

As in Experiment 1, participants were students enrolled in an online Brain and Behavior course. Unlike Experiment 1, this course took place in a summer semester, covering the material in eight rather than 15 weeks. Thirty-two students enrolled in the class. Data from 27 students (those who took tests in Units 1 and 2 and who missed no more than one quiz out of three in either unit) were used in analyses (two students whose data were included in the analyses eventually dropped the course). Other design elements remained unchanged from Experiment 1.

2.1.2. Procedure

The quiz materials were the same as in Experiment 1. Procedures differed from Experiment 1 primarily as required by the differences between a school-year and summer course. Rather than covering a unit every three weeks, an entire unit could be covered in one-and-a-half weeks.

(multiple choice and short answer either identical to or related to the exam questions), reflected performance on 14 questions (seven per unit). Seven participants missed one quiz—exposure to 14 questions—in either Unit 1 or 2; two of these participants missed one quiz (or read statements) in each unit. Accordingly, for these subjects, information from the missed quizzes (or read statements) was included in the no-quiz (control) scores for these participants.

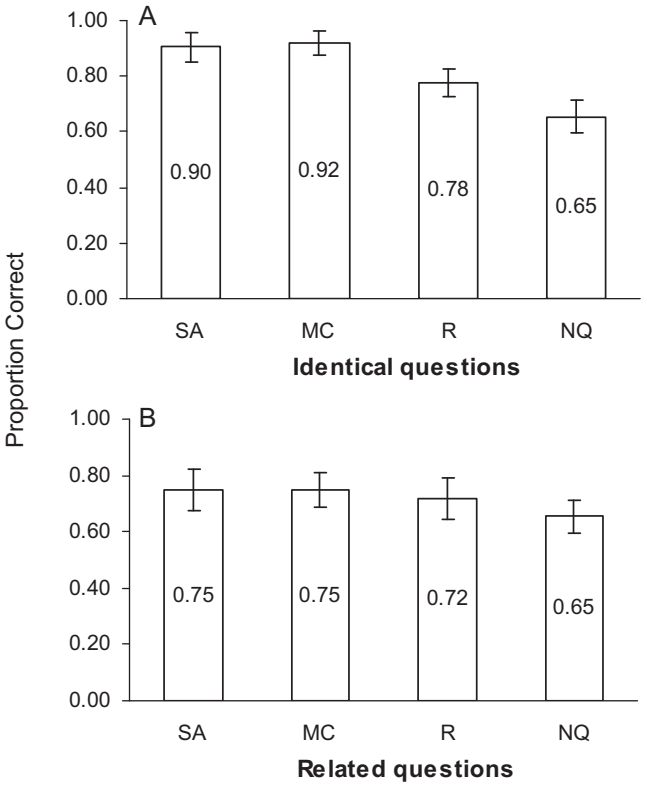

Fig. 2. Mean proportion correct on the unit tests as a function of quiz exposure condition and unit-test question type in Experiment 2 (error bars represent 95% confidence intervals). SA: short answer; MC: multiple choice; R: read; NQ: no quiz.

2.2. Results and discussion

2.2.1. Quiz performance

A within-subjects ANOVA revealed no significant difference in the number of times each type of quiz was taken ($F < 1$; short answer $M = 4.43$, SD = 2.04; multiple choice $M = 4.26$, SD = 53; read-only $M = 4.15$, SD = .36).

As expected, quiz performances significantly improved across repeated quiz attempts (see Table 2 for means), $F(3, 78) = 210.11$, $p < .001$, $\eta_p^2 = .89$. Quiz performance was higher on multiple choice quizzes than on short answer quizzes primarily for initial quiz attempts, $F(3, 78) = 30.41$, $p < .001$, $\eta_p^2 = .55$, for the quiz type and quiz attempt interaction.

2.2.2. Unit exam performance

As in Experiment 1, the proportion of correct unit test answers was the dependent variable.[2] Consistent with the previous experiment, performance on exam items was significantly better when preceded by identical quiz questions in either short answer format or multiple choice format relative to the no-quiz control items, $t(26) = 10.18$, $p < .001$, $d = 1.80$, and $t(26) = 10.42$, $p < .001$, $d = 2.08$, respectively (see Fig. 2, Panel A). Also, both short answer quizzing and multiple choice quizzing (using questions identical to the exam items) produced a significant benefit over presenting the identical target content for reading (additional study), $t(26) = 4.25$, $p < .001$, $d = 1.00$, and $t(26) = 5.74$, $p < .001$, $d = 1.23$, respectively. It is noteworthy that (1) the quizzing benefits here and in Experiment 1 reflect large effect sizes and (2) the quizzing benefits were obtained relative to the read condition even though reading of question answers also produced significant benefits relative to the no-exposure condition, $t(26) = 4.64$, $p < .01$, $d = .89$.

Also replicating Experiment 1, when quiz items were related to the exam items, short answer quizzes and multiple choice quizzes enhanced performance on the exam relative to no-quiz items, $t(26) = 3.18$, $p < .01$, $d = .60$, and $t(26) = 4.56$, $p < .001$, $d = .62$, respectively (see Fig. 2, Panel B). By contrast, reading the related content did not significantly benefit final exam performance relative to the no-quiz control items ($t(26) = 1.72$). As before, the related quiz questions (in either short-answer or multiple-choice formats) did not produce significantly higher performance than reading information related to the exam questions (both ts < 1).

3. General discussion

This study represents the first well-controlled experimental demonstration of an advantage of on-line quizzing (even relative to additional study, i.e., the read condition, in a college course using actual course content, thereby establishing an important extension of laboratory testing effects into the classroom setting (see Glass et al., 2008, for an experiment with on-line and in-class quizzes but without a study control). Several results are noteworthy. First, repeated quizzing of questions identical to those on the exams substantially enhanced exam performance relative to content that was not re-presented (control content) and content that was re-presented for reading (study). The magnitude of this benefit was impressive in terms of both the large effect sizes obtained and in terms of the increase in projected grades. In Experiment 2, for instance, same-stem quizzing improved unit exam performance to levels of 90% or better (grade of A) from that of 78% observed when the same content was presented for reading (grade of C).

Second, repeated quizzing of questions that were related to those on the exams but were focused on different aspects of the content than queried on the exam, along with elaborative feedback, produced consistent improvement on exam performance relative to the control items, with effect sizes being of medium magnitude. This experimental finding reinforces the correlational reports that repeated quizzing can enhance performance on exams for which questions are not identical to those on the quizzes (Angus & Watson, 2009; Kibble, 2007; see also related experimental findings of Glass, 2009).

Third, the finding that repeated multiple-choice quizzing was not less effective than short-answer quizzing qualifies previous experimental work reporting that when a single quiz with feedback is administered, multiple choice quizzing is not as effective as short-answer quizzing, even when the final test is also multiple choice (Kang et al., 2007, Experiment 2; McDaniel et al., 2007). The high levels of learning (possibly overlearning) of the targeted content fostered by repeated quizzing (note performances approaching ceiling by the 3rd quiz; Table 2) appears to obviate advantages of question formats requiring effortful retrieval (e.g., recall) relative to formats for which retrieval is relatively less effortful but still required (e.g., recognition; see Kang et al., 2007, Experiment 1). Possibly, repeated multiple-choice quizzing gains advantage (relative to a single quiz) because of the repeated retrieval required (Karpicke & Roediger, 2007), though the open-book option leaves this interpretation uncertain. Regardless, the practical implication is that with repeated quizzing, a multiple-choice format can be as effective as a short-answer format, thereby allowing the pragmatic benefit of automated grading without sacrificing the potency of the testing effect.

Several interpretations of these benefits that are unrelated to the quizzing process can be ruled out by the advantage of the quizzing (identical item) conditions compared to the reading condition. First,

[2] Scores were calculated as in Experiment 1. One student missed a short answer quiz.

this quizzing effect was not a consequence of additional exposure per se of target content. Note that students accessed the reading representations as frequently as they accessed the quizzes (as they were instructed to do). Accordingly, the content was exposed as frequently for read re-representations as for quizzes. Second, the present quizzing effects are unlikely to reflect study guide usage (cf. Daniel & Broida, 2004) because the reading condition would have also provided an effective study guide. By contrast, previous correlational and quasi-experimental demonstrations of on-line quizzing, as well as in-class quizzing, outcomes are open to these above interpretations (e.g., Angus & Watson, 2009; Kibble, 2007; Lyle & Crawford, 2011).

The present experiments thus provide some of the strongest evidence to date that taking quizzes with feedback can positively affect student learning outcomes in college-level courses (see also Glass, 2009, though with no reading control), and even when the quizzes are on-line and unsupervised. From a theoretical perspective, there are several candidate interpretations of the present benefits of quizzing. Though the present study was not designed to disentangle these interpretations, we mention the more prominent possibilities.

One possibility is that quizzing serves a formative assessment function, identifying for students content that is not yet well learned and that merits further study. More specifically, the current feedback procedure, which provided modest elaboration on each quizzed item, allowed students to directly access correct answer information to gauge their response accuracy and to support additional study of that information. This effect of feedback is suggested by the consistent enhancement produced by the related-quiz conditions (relative to the no-quiz control). The feedback included additional information that could serve as hints (or even the answer) to the related questions on the unit exams (see Appendix A), information that was not directly needed for answering the quiz item itself. In the reading presentation condition, the same "feedback" was available, and although unit exam performance for target information in the related-read condition was not significantly lower than for the related quizzed items, the related read condition did not significantly improve unit exam performance over the no-quiz control. These nominal patterns may suggest that learners focus more attention or recruit more effective encoding strategies when target information serves as corrective feedback (see Bahrick & Hall, 2005). More generally, the provision of feedback likely augmented the overall benefits of quizzing (Butler & Roediger, 2008; McDaniel & Fisher, 1991).

Another idea is that students attempted to answer the quizzes without consulting their text or notes, and the retrieval of the requisite information promoted good retention (McDaniel & Masson, 1985; Roediger & Karpicke, 2006b). In light of the enhanced performance produced by the related quiz items, retrieval processes possibly also activated related information to that targeted by the initial test item (see Chan et al., 2006, for laboratory support of this possibility; but see discussion above). Students' relatively poor performance on their first quiz attempts for both multiple choice and short-answer quiz items supports the idea that many (and perhaps all of the students) approached the first quizzes as closed book (if not, then much higher performance would be expected; cf. Kibble, 2007). On subsequent quizzes students may have consulted the text (or notes), as performance was much higher (or this could indicate learning from the initial quiz attempt, either from the feedback or from further study). If so, then the results suggest that a combination of attempted retrieval of quiz answers both with closed and open book can promote learning just as closed book quizzes can (see also Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008), and thus allowing open-book access for on-line (and perhaps in-class) quizzing does not necessarily undermine the benefits of quizzing (cf. Daniel & Broida, 2004).

To conclude, in most educational contexts target information must be mastered. In some of these contexts, such as physics and mathematics courses, the numerous problems that students solve to prepare for exams may provide sufficient practice; accordingly, in these contexts explicit quizzing may not be necessary (but see Lyle & Crawford, 2011, for quizzing effects in a college statistics course). For the many college courses that do not afford problem-solving exercises, the present results suggest that quizzing along with feedback can be an effective technique to assist mastery and retention of content. Moreover, we note that our results are likely not limited to on-line quizzing and web-based courses. Recent studies using in-class quizzes and feedback (and in some cases with clicker responses), have also reported that quizzing improves summative test performances (see Lyle & Crawford; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011, in middle school classrooms; Glass, 2009; Glass et al., 2008, using a mix of online and in-class quizzes in college courses). Our findings clearly indicate that quizzes with questions identical to those on the exam can provide robust gains on unit exams relative to reading or studying the content (read condition). Thus, identical quizzes provided practice that was not promoted by the read condition, possibly including practice at retrieving responses. Also, learning the responses might have been potentiated at least in part by providing correct answer feedback after the quiz items (see Butler & Roediger, 2008). Regardless, on the view that "factual knowledge must precede skill" (Willingham, 2009, p. 19), and more particularly that retrieval practice facilitates fluent recall of facts so as to free up students' cognitive resources for higher-order processes such as critical thinking and analysis (Willingham), quizzing may contribute to aspects of learning that scaffold higher-order understanding of a content domain.

Appendix A.

Samples of quiz[1] and exam materials used in Experiments 1 and 2

Multiple Choice Quiz Questions[2]	Feedback	Related Exam Questions[3]
With functions regulated by an overlapping set of structures, humans retain some behavioral skills in lower levels such as spinal cord in common with _____. a. inhibitory brain centers b. autonomic ganglia c. single-celled organisms d. primitive vertebrates	With functions regulated by an overlapping set of structures, humans retain some behavioral skills in lower levels such as spinal cord in common with primitive vertebrates.	Humans have behavioral skills in lower levels such as spinal cord _____. a. only if myelination proceeds into adulthood b. as a result of extreme training during childhood c. if they are raised in impoverished environments d. that overlap with more sophisticated controls in higher centers
Information coming INTO a structure (arriving) is called _____ . a. different b. efferent c. superior d. afferent	Information coming INTO a structure (arriving) is called afferent. Information leaving a structure (exiting) is called efferent.	Information leaving a nervous system structure is called _____. a. different b. affective c. superior d. efferent
Tissue covering the brain and spinal cord can become infected, a condition known as _____. a. encephalitis b. caudal c. efferent d. meningitis	Tissue covering the brain and spinal cord (the meninges) can become infected, a condition known as meningitis.	The 3 layers of tissue covering the brain and spinal cord (dura, arachnoid layer, and pia mater) are known as the _____. a. blood brain barrier b. central ventricles c. lobes d. meninges
Cells in the cerebral cortex are organized into _____ . a. nerves	Cells in the cerebral cortex are organized into six layers. In different cortical	Cells in the human _____ are organized into six layers, and different Brodmann's areas have distinctive composition of these

Appendix A (*Continued*)

b. pyramids c. ventricles d. six layers	regions, known as Brodmann's areas, the exact layering is distinctive.	layers. a. limbic cortex b. adrenal cortex c. ventricles d. neocortex
Keeping the bodily organs alive requires functioning of _____ structures. a. frontal lobe b. ventricular c. rostral d. brainstem	Keeping the bodily organs alive requires functioning of brainstem structures. In some animals, this constitutes virtually their entire brain.	Human brainstem contains structures that _____ regulating respiration and circulation. a. have been replaced by neocortex for b. produce stem cells for c. can be removed surgically to aid in d. keep body organs alive by
Although we have _____ symmetry, some functions of the body are controlled by just one structure that is on one side of the brain. a. bilingual b. melodic c. segmental d. bilateral	Although we have bilateral symmetry, some functions of the body are controlled by just one structure that is on one side of the brain. Language is usually processed in the left hemisphere and spatial processing is usually on the right.	Language is usually processed in the left _____ and spatial processing is usually on the right. a. parietal lobe b. medulla c. thalamic nucleus d. cerebral hemisphere

[1]The short-answer quiz questions used the same stem as the multiple-choice quiz questions.
[2]Correct answers are all (d), but in the experiments all lures were randomized.
[3]The identical exam questions were exactly like the multiple choice quiz questions presented in the first column.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L. & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, *22*, 861–876.

Angus, S. D. & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology*, *40*, 255–272.

Bahrick, H. A. & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, *52*, 566–577.

Butler, A. C. & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*, 604–616.

Callender, A. A. & McDaniel, M. A. (2007). The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology*, *99*, 339–348.

Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642.

Chan, J. C. K., McDermott, K. B. & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553–571.

Daniel, D. D. & Broida, J. (2004). Using web-based quizzing to improve exam performance: Lessons learned. *Teaching of Psychology*, *31*, 207–208.

Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, *29*, 831–848.

Glass, A. L., Brill, G. & Ingate, M. (2008). Combined online and in-class pretesting improves exam performance in general psychology. *Educational Psychology*, *28*, 483–503.

Glover, J. A. (1989). The 'testing' phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*, 392–399.

Kang, S. H. K., McDermott, K. B. & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528–558.

Karpicke, J. D. & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162.

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *Advances in Physiology Education*, *31*, 253–260.

Kolb, B. & Whishaw, I. Q. (2006). *An introduction to brain and behavior* (2nd ed.). New York: Worth Publishers.

Lyle, K. B. & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, *38*, 94–97.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*, 51–57.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B. & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*, 399–414.

McDaniel, M. A., Anderson, J. L., Derbish, M. H. & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494–513.

McDaniel, M. A. & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, *16*, 192–201.

McDaniel, M. A. & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371–385.

Nungester, R. J. & Duchastel, P. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, *74*, 18–22.

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A. & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term benefits from quizzing. *Journal of Experimental Psychology: Applied*, doi:10.1037/a0026252 (Advanced online publication)

Roediger, H. L., III & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H. L., III & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Roediger, H. L., III & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1155–1159.

Willingham, D. T. (2009). *Why don't students like school?* San Francisco, CA: Jossey-Bass.