MEMORY, 2011, 19 (8), 836–852

Psychology Press
Taylor & Francis Group

Expectancy of an open-book test decreases performance on a delayed closed-book test

Pooja K. Agarwal and Henry L. Roediger III

Washington University in St. Louis, St. Louis, MO, USA

Two experiments examined the influence of practice with, and the expectancy of, open-book tests (students viewed studied material while taking the test) versus closed-book tests (students completed the test without viewing the studied material) on delayed retention and transfer. Using GRE materials specifically designed for open-book testing, participants studied passages and then took initial open- or closed-book tests. Open-book testing led to better initial performance than closed-book testing, but on a delayed criterial (closed-book) test both types of testing produced similar retention after a two-day delay in Experiment 1. In Experiment 2 participants were informed in advance about the type of delayed criterial test to expect (open- or closed-book). Expecting an open-book test (relative to a closed-book test) decreased participants' time spent studying and their delayed test performance on closed-book comprehension and transfer tests, demonstrating that test expectancy can influence long-term learning. Expectancy of open-book tests may impair long-term retention and transfer compared to closed-book tests, despite superior initial performance on open-book tests and students' preference for open-book tests.

Keywords: Open-book tests; Testing effect; Test expectancy; Feedback; Transfer.

In classroom settings, tests and quizzes are typically administered for assessment purposes. Laboratory and applied research, however, demonstrate that tests and quizzes not only measure knowledge, but also change and enhance retention of information (e.g., Roediger & Karpicke, 2006a). This *testing effect* has been well established, and recent research has focused on educational implications of the testing effect (for reviews, see Roediger, Agarwal, Kang, & Marsh, 2010; Roediger & Karpicke, 2006b). In addition, prior research has evaluated the effect of initial and final test format on criterial performance, for instance free recall versus cued recall versus recognition tests, and also multiple-choice versus short answer tests (Carpenter & DeLosh, 2006; Duchastel & Nungester, 1982; Glover, 1989; Kang, McDermott, & Roediger, 2007).

Another important distinction exists between test or quiz formats commonly found in educational settings: closed-book and open-book. During a closed-book test students are not allowed to refer to notes or textbook materials. During an open-book test, however, students are allowed to refer to notes or textbook materials. Although these two types of tests are usually used for

Address correspondence to: Pooja K. Agarwal, Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130, USA. E-mail: pooja.agarwal@wustl.edu

This research was supported by the National Science Foundation Graduate Research Fellowship Program and the Harry S. Truman Scholarship Foundation (awarded to the first author), and the James S. McDonnell Foundation 21st Century Science Initiative grant, Applying Cognitive Psychology to Enhance Educational Practice: Bridging Brain, Mind, and Behavior Collaborative Award (awarded to the second author). We thank Jake Sanches, Emily Rosenzweig, and Jane McConnell for their assistance. We also thank David Balota, Kathleen McDermott, and Andrew Butler for valuable discussions, and Yana Weinstein for providing helpful comments on a draft of this manuscript. This project was completed as part of a master's thesis by the first author.

http://www.psypress.com/memory http://dx.doi.org/10.1080/09658211.2011.613840

assessment purposes, we can ask a similar question as asked in previous research: which test format, closed-book or open-book, is most effective in enhancing learning?

In the first systematic study of the testing effect with closed-book and open-book tests, Agarwal, Karpicke, Kang, Roediger, and McDermott (2008) had students read a series of passages, each of which was followed by a closed-book test, a closed-book test with feedback (where participants graded their own responses using the passage after testing), or an open-book test (where participants had access to the passage during testing). Students also read a passage in a study-only condition, which was not followed by an initial test. After 1 week students returned for a second session and completed closed-book tests over each passage studied during the first session. Across two experiments Agarwal et al. found that although initial test performance was highest in the open-book test condition, the open-book test and closed-book test with feedback conditions resulted in similar delayed performance after 1 week, and delayed performance following all three initial test conditions was greater than performance following the study-only condition (i.e., a significant testing effect occurred). However, greater forgetting from the first to second tests occurred when the initial test was open-book.

One possible criticism of the Agarwal et al. (2008) experiment is that the educational passages and fact-based short answer questions used were not appropriate for open-book tests, because open-book tests are supposed to enable a student to integrate and transfer information (Jacobs & Chase, 1992). Instructors maintain that open-book tests are designed to require students to apply knowledge, rather than memorise or restate it; therefore if a student uses higher-order thinking skills during an initial open-book test, benefits for delayed retention and transfer on final criterial tests may emerge (Feller, 1994; Theophilides & Koutselini, 2000). We used such materials in the present research.

A second possible criticism of Agarwal et al. (2008) is that students did not know which type of initial test to expect after studying: an open-book test or a closed-book test. Consider, however, that even instructors who support the use of open-book tests acknowledge that students may not find open-book tests to be as challenging as closed-book tests and that students will often spend less time studying for open-book tests (Eilertsen & Valdermo, 2000; Ioannidou, 1997; Jacobs & Chase, 1992). We examined the issue of study time in the current experiments, too.

The present experiments were aimed at building on the work of Agarwal et al. (2008) by (1) determining possible differences in memorial benefits between open-book and closed-book tests using materials specifically designed for open-book testing that involved comprehension and transfer questions requiring integration across a passage, and (2) examining the effect of test expectancy on time on task during studying, open-book tests, and closed-book tests, in order to evaluate whether students spend less time studying for and completing open-book tests in comparison to studying for and completing closed-book tests.

The passages and tests used in the present experiments were drawn from a Graduate Record Examination (GRE) test preparation book (Research & Education Association, Inc., 2008). Comprehension questions from the verbal section of the GRE require students to analyse relationships, apply ideas to novel situations, and draw inferences (Educational Testing Service, 2002). During typical GRE test taking students are both allowed and required to refer back to the passage at hand, a process akin to an open-book test in the classroom.

Final criterial tests in the current experiments comprised both original GRE comprehension questions and higher-order transfer questions that required participants to indicate why a certain detail from the passage was true. The answer required for an initial comprehension item was embedded in the question stem of the final transfer item; however, the answer for the transfer item (a causal reason for why a detail was true, also known as a causal antecedent) was not previously quizzed but could be inferred from the passage (Graesser, Singer, & Trabasso, 1994). For example, a GRE comprehension question from a passage about William Penn and the colonisation of Pennsylvania included the question:

> Which of the following statements would the author most likely agree with?
>
> (A) The King of England imposed severe restrictions on Penn's land grant
> (B) Penn was an opportunistic businessman
> (C) The Indians of Pennsylvania were savages
> (D) Penn was too friendly with the King of England

(E) Indians didn't bother the settlers because they were permitted to practice their own religion

The correct answer for the comprehension question, in the context of the studied passage, was B. For the final transfer test given later, a corresponding question written by the experimenter asked the following:

Why was Penn an opportunistic businessman?

(A) Because he made a personal fortune while governing Pennsylvania
(B) Because he purchased Pennsylvania for much less than it was worth
(C) Because he sold off his land quickly enough to make large profits
(D) Because he became wealthy while using the King's money
(E) Because he taxed all of the successful businesses

The correct answer for the transfer question was A. Notice that the stem of the transfer question incorporated the correct answer from the earlier comprehension question.

In order to elucidate the distinctions between the two types of questions (comprehension and transfer), we call the reader's attention to Barnett and Ceci's (2002) taxonomy of transfer, as well as Bloom's taxonomy (Krathwohl, 2002). Barnett and Ceci included two main factors or areas in which transfer can occur: content (what is transferred) and context (when and where transfer occurs). The types of comprehension and transfer questions used in the two experiments reported here fall under the ''memory demand'' content factor and the ''knowledge domain'' context factor. Memory demand questions refer to items in which a fact or concept must be retrieved, whereas the knowledge domain assessment requires that the fact be applied in a new context.

Regarding the transfer of a memory demand, students initially answered comprehension questions based on specific ideas from the passage (memory demand) and then answered transfer questions that required causal understanding and reasoning (knowledge demand). The present procedure requires recall of a learned fact and its applicability, as well as the ability to execute the required memory demand and transfer it to a new task (Barnett & Ceci, 2002). Regarding transfer of context within the knowledge domain, Barnett and Ceci described the knowledge domain as ''the knowledge base to which the skill is to be applied'' (p. 623). Because the comprehension and transfer questions in the present study tested knowledge about the same passage details, any transfer within the knowledge domain would be considered near transfer. Thus the transfer of memory demand from comprehension questions to transfer questions is more critical in our experiments than the transfer of knowledge between domains. In addition, the comprehension questions in the current study fit within Bloom's comprehension category, whereas the transfer questions fit within the analysis category (Krathwohl, 2002).

By deliberately using two different types of questions, detailed GRE comprehension questions and conceptual/inferential transfer questions, we hoped to evaluate any potential benefits of open-book testing that did not emerge with Agarwal et al.'s (2008) fact-based materials. We hoped to show that tests do not simply improve learning of facts but also the application and transfer of those facts in new contexts (see Butler, 2010), while also examining how test expectancy would influence delayed performance. Performance on delayed comprehension and transfer questions may be greater following an initial open-book test than an initial closed-book test, because open-book testing may engage higher-order cognitive skills when using appropriate materials. On the other hand, performance on delayed tests may be reduced following an initial open-book test relative to an initial closed-book test, if students study less effectively when expecting open-book tests.

In Experiment 1 we aimed to extend Agarwal et al.'s (2008) research by using three of their key conditions (study-only, closed-book test with feedback, and open-book test), while using materials appropriate for open-book tests and measuring final criterial performance on both comprehension and transfer questions (in a closed-book procedure) after a two-day delay. It is important to note that additional conditions from Agarwal et al. (2008), such as restudying conditions (controlling for exposure time) and a closed-book test (without feedback) condition, were not included in the present study because controlling for time on task and examining benefits from feedback versus no feedback were not of primary interest in the current study (see Agarwal et al., 2008, for the relevant conditions and discussion). Instead, including the study-only

condition in the present study allowed us to measure a basic testing effect (defined as comparing performance following study of a passage plus an initial test to performance following studying a passage once; Wheeler & Roediger, 1992). In addition, including a closed-book test with feedback condition in the present study allowed us to compare a traditional open-book test condition (in which subjects receive feedback from the passage *during* an initial test) to a similar condition in which participants receive feedback from the passage, but *after* completing an initial closed-book test.

EXPERIMENT 1

Method

Participants. A total of 72 participants (M_{age} = 19.5 years old, 46 females) were recruited from the Department of Psychology human subject pool. Participants received either credit towards completion of a research participation requirement or cash payment.

Design. Three within-participant initial learning conditions (study-only, closed-book test with feedback, open-book test) were crossed with two delayed test types (comprehension and transfer), for a 3×2 within-participant design. Six passages were presented in the same order for all participants, but the order in which the three initial learning conditions and two delayed test types occurred was counterbalanced using a Latin Square, creating six counterbalancing orders. Twelve participants were randomly assigned to each of the six orders, and all conditions appeared once in every ordinal position an equal number of times across passages and participants.

Materials. Six passages, averaging 438 words in length, were adapted from a GRE test preparation book (Research & Education Association, Inc., 2008). The six passages (''Plant Adaptations'', ''Robert Goddard'', ''Submarines'', ''William Penn'', ''Taxonomy'', and ''Michael Faraday'') covered scientific or biographical topics. Initial multiple-choice comprehension questions were adapted from the same test preparation book, whereas the experimenter created final multiple-choice transfer questions such that participants were asked to choose ''why'' a particular idea that was stated in the passage and quizzed on the initial comprehension test was true. All questions on comprehension tests had a corresponding ''why'' question on the transfer tests. All tests were composed of six multiple-choice questions with five-alternative responses.

Participants completed comprehension tests during the first session of the experiment. During the second session two days later, participants completed closed-book comprehension and transfer tests in alternating order (depending on the counterbalancing scheme). Comprehension questions encountered in Session 2 were identical to those in Session 1; however, the five response options were presented in a new random order for the second session. All passages and tests were studied and completed in paper-and-pencil format, and the order of questions for each test was the same across all participants.

Procedure. Participants were tested individually or in small groups. All passages and tests were pre-arranged (according to the counterbalancing order) in one blue folder for each participant, always face down so that participants could not view the next set of passages or test questions until prompted (i.e., the first passage to be read was face down, but at the top of the stack). An empty red folder was also provided for each participant for completed passages and tests. Each participants was seated at a computer and used an E-Prime 1.0 program (Schneider, Eschman, & Zuccoloto, 2002) that provided instructions and recorded time spent on each phase of the experiment. The experimenter was outside of the testing room and monitored compliance with all instructions via a window. All study and test periods were self-paced.

In Session 1 participants were instructed that they would read several passages, which might or might not be followed by a test. During a study period instructions directed participants to take a passage from the blue folder, read it at their own pace, place it face down in the red folder when they were finished, and push spacebar on the keyboard to move on to the next set of instructions (the E-Prime program recorded time spent while studying). During a test participants were asked to take the corresponding test from the same blue folder (which would now be face down at the top of the stack, according to counterbalancing order), circle a multiple-choice alternative for every question, place the test face down in the red folder when they completed the test, and push the spacebar for the next set of

instructions (the E-Prime program recorded time taken during testing). Participants could not move through the folder of materials or the instructions on the computer in advance; thus participants did not know whether to anticipate a closed-book test, an open-book test, or another passage throughout Session 1.

During Session 1 participants read six passages: two in the study-only condition, two in the closed-book test with feedback condition, and two in the open-book test condition. Once participants completed one condition (e.g., studying and taking an open-book test), they moved on to the next condition, according to their counterbalancing order. In the study-only condition participants read the passage one time and were not tested on it; instructions directed participants to move on to the next passage. In the closed-book test with feedback condition participants read the passage, completed the multiple-choice comprehension test without viewing the passage, and then they were asked to take the corresponding passage out of the red folder and check their answers (to mimic the type of ongoing feedback received in the open-book test condition). Specifically, participants were instructed to write ''correct'' next to responses they believed were correct (based on information from the passage; participants were not informed of the actual correct and incorrect answers) and to write ''incorrect'' next to responses they believed were incorrect, without changing their original answers. In the open-book test condition participants read the passage one time and then were able to view the passage while completing the multiple-choice comprehension test.

Session 2 occurred two days after Session 1. Participants completed multiple-choice tests over all six passages without restudying or reviewing the passages (i.e., the final tests were closed-book); three were comprehension tests repeated from Session 1 and three were transfer tests (one for each initial learning condition), in alternating order (according to requirements of the counterbalancing scheme). The entire procedure lasted approximately 90 minutes across the two sessions. At the end of the experiment, participants were debriefed and thanked for their time.

Results

All results in the current experiments were significant at an alpha level of .05 unless otherwise noted.

Initial test performance. Initial comprehension test performance is shown in the first column of Table 1. As expected, initial test performance was significantly greater on open-book tests ($M = .69$) in comparison to closed-book tests ($M = .60$), $F(1, 71) = 13.34$, $\eta_p^2 = 0.16$. Of course this difference might be caused by performance in the open-book test condition being measured *while* feedback (i.e., referring to the passage) was accessible, whereas performance in the closed-book test condition was measured *before* feedback was accessible. Differences in initial performance do not imply that participants necessarily used different study or retrieval strategies during the two tests; rather, differences in performance may have been a result of the timing of feedback and access to the passage.

After completing the closed-book test, participants received feedback on each question by grading their own answers as correct or incorrect, and participants accurately self-graded 77% of items in this condition. Specifically, participants wrote ''correct'' next to responses that were actually correct and wrote ''incorrect'' next to responses that were actually incorrect on 661 of 864 possible items (6 items per 2 passages per 72 participants), indicating that participants were

TABLE 1
Initial and delayed test performance (proportion correct) in Experiment 1

	Initial Comprehension Test	*Delayed Comprehension Test*	*Delayed Transfer Test*	*Delayed Average*
Study-only		.49 (.03)	.60 (.03)	.55 (.02)
Closed-book test with feedback	.60 (.02)	.61 (.03)	.70 (.03)	.66 (.02)
Open-book test	.69 (.02)	.63 (.03)	.70 (.03)	.67 (.02)
Average	.65 (.02)	.58 (.02)	.67 (.02)	

Delayed tests occurred two days after initial comprehension tests. Standard errors of the mean are displayed in parentheses, $n = 72$.

processing feedback during the self-grading process (even if far from perfectly; see also Rawson & Dunlosky, 2007).

Delayed test performance. Performance on the two-day delayed test is shown in Table 1. Two separate one-way analyses of variance (ANOVAs) on learning condition (study-only, closed-book, open-book) were conducted for each of the delayed test types (comprehension, transfer). Considering first comprehension test performance, a main effect existed such that retention was greatest for the open-book test condition ($M = .63$), followed by the closed-book test with feedback ($M = .61$) and study-only ($M = .49$) conditions, $F(2, 142) = 9.24$, $\eta_p^2 = 0.12$. Comprehension performance for the open-book test condition was significantly greater than performance for the study-only condition, $t(71) = 3.80$, $d = 0.58$, and comprehension performance for the closed-book test with feedback condition was also greater than the study-only condition, $t(71) = 3.98$, $d = 0.54$. The present results demonstrate the memorial benefit of testing compared to studying on comprehension tests, regardless of the type of initial test. Although the open-book test condition resulted in slightly greater comprehension performance than the closed-book test with feedback condition, this difference was not significant, $t < 1$.

Delayed transfer test performance was similar for the open-book and closed-book test with feedback conditions ($M = .70$ for both conditions), followed by performance for the study-only condition ($M = .60$), resulting in a significant main effect of learning condition on delayed transfer test performance, $F(2, 142) = 5.70$, $\eta_p^2 = 0.07$. Again, performance for open-book and closed-book test conditions were significantly greater than performance for the study-only condition, $t(71) = 2.85$, $d = 0.42$, and $t(71) = 2.79$, $d = 0.39$, respectively. The results confirm the robust effects of initial tests on the transfer of knowledge (see Butler, 2010), as well as initial tests on delayed retention (Agarwal et al., 2008; Kang et al., 2007).

Time on task. Time on task during study periods (reading time) and during the initial and delayed tests is shown in Table 2. Recall that time spent studying and testing was recorded using E-Prime. A one-way ANOVA on reading times by learning condition revealed no significant differences across the study-only ($M = 144$ s), open-book ($M = 145$ s), or closed-book ($M = 148$ s) learning conditions, as would be expected because participants did not know what condition to expect before reading the passage, $F < 1$, $p = .737$. Time spent completing initial comprehension tests was much greater when participants took open-book tests ($M = 254$ sec) in comparison to closed-book tests ($M = 157$ sec), $F(1, 71) = 106.99$, $\eta_p^2 = 0.60$, suggesting that participants made use of the available passage while completing the open-book test. Participants spent 173 seconds self-grading their test in the closed-book test with feedback condition, and total time spent testing and processing feedback in the closed-book condition ($M = 331$ sec) was significantly greater than time spent testing (and presumably processing feedback while referring to the passage) in the open-book condition ($M = 254$ sec), $F(1, 71) = 65.78$, $\eta_p^2 = 0.48$.

For the second session, two separate one-way ANOVAs on test completion times by learning condition (study-only, closed-book, open-book) were conducted for each of the delayed test types (comprehension, transfer). For the comprehension test, time spent for the study-only condition on the delayed test ($M = 145$ s) was greater than time spent for the closed-book ($M = 96$ s) and open-book ($M = 100$ s) conditions, $F(2, 142) = 24.70$, $\eta_p^2 = 0.26$. However, completion times on the delayed transfer test did not differ significantly across the three learning conditions (times ranging from 108 to 113 s), $F < 1$, $p = .695$. Presumably, because participants had completed the comprehension questions previously in the two testing conditions, they were able to answer them more quickly on the delayed test relative to the study-only condition. Participants had not taken the transfer tests previously and so no differences were expected, nor were any observed.

Discussion

Similar to the results from Agarwal et al. (2008), significant testing effects were found such that delayed comprehension and transfer test performance following initial open-book and closed-book tests (with feedback) was greater than delayed test performance for the study-only condition. As in previous experiments, potential benefits from initial open-book tests did not emerge on either the delayed comprehension tests or the delayed transfer tests, even when using materials (GRE passages and tests) specifically designed for open-book testing.

TABLE 2
Time on task (seconds) in Experiment 1

	Session 1			*Session 2*	
	Reading	*Initial Comprehension Test*	*Self-Grading*	*Delayed Comprehension Test*	*Delayed Transfer Test*
Study-only	144.2			145.1	108.4
Closed-book test with feedback	148.1	157.2	173.7	95.9	113.1
Open-book test	145.2	254.0		100.1	109.4

Session 2 occurred two days after Session 1, $n = 72$.

Again, delayed performance was similar for the open-book and closed-book test with feedback conditions, indicating that although open-book tests yield higher initial performance, this boost does not persist over a two-day delay using appropriate materials (consistent when using different materials over a week-long delay in Agarwal et al., 2008).

Surprisingly, performance on transfer questions ($M = .67$, collapsed over learning conditions) was greater than performance on comprehension questions ($M = .58$, collapsed over learning conditions). This result was puzzling, considering that the transfer questions were intended to require inferential (and possibly more challenging) retrieval processes and the comprehension tests were simply repetitions of earlier tests. Of course, it is difficult to make tests that are of equivalent difficulty, so it seems we simply constructed transfer questions that were easy. Recognition of causal reasons (assessed by the transfer questions) may be easier than recognition of specific ideas from the passages (assessed by the comprehension questions). Thus higher performance on multiple-choice transfer questions than on multiple-choice comprehension questions (across all three initial learning conditions) might be the result of such item differences. The important point is that we found testing effects on both forms of the delayed criterial test.

Finally, although participants spent an additional minute during the closed-book test with feedback condition in the first session, delayed test performance for the closed-book test with feedback and open-book test conditions was similar. Thus open-book testing may be more efficient than closed-book testing with feedback in promoting later learning, where efficiency is defined as spending the least amount of time in order to learn information well enough to recall it later (Pyc & Rawson, 2007). However, less is not always more—the learning condition in which participants spent the least amount of time during the first session, the study-only condition, also produced the lowest performance (in terms of delayed comprehension and transfer test performance) for the second session. Thus testing and processing feedback simultaneously may take less time than testing and processing feedback consecutively, although the 1-minute processing advantage for the open-book test condition did not result in a discernable increase in delayed performance. Another possibility is that, in the open-book condition, participants may have consulted the passage only when unsure of an answer, whereas in the closed-book condition participants were instructed to score *every* answer, whether correct or incorrect. Although it has been demonstrated that feedback on correct answers is often beneficial for retention (Butler, Karpicke, & Roediger, 2008), the checking procedure may have added to participants' time on task in the closed-book test condition when they would not ordinarily check their confident or correct answers.

EXPERIMENT 2

The current results from Experiment 1 provide additional evidence that testing effects can be obtained with both open-book and closed-book tests; they also replicate findings from Agarwal et al. (2008) that initial open-book tests and closed-book tests with feedback do not produce different levels of performance on delayed retention tests, even when using materials designed for open-book tests. One point of departure between previous experiments and typical educational practice is that students usually know the type of quiz or exam to expect with regards to open-book versus closed-book format. In Agarwal

et al. (2008), and in Experiment 1 of the current study, participants were unaware of the type of initial or final test to expect and thus could not prepare in the manner they usually would when they have such foreknowledge. Participants might have studied the passages in preparation for closed-book tests, because closed-book tests are more prevalent in both classroom and research settings (Feldhusen, 1961; Theophilides & Koutselini, 2000). If test expectancy affects students' study procedures and performance on open- and closed-book tests, then perhaps we should not be surprised by the similar delayed test performance in prior research and in Experiment 1.

In Experiment 2 we told participants which type of test to expect, to try to capture their natural study strategies for open- and closed-book tests. We included two design components intended to address methodological concerns articulated in the test expectancy literature (Neely & Balota, 1982; Schmidt, 1980, 1983). First, participants received four initial passages and practice tests, two closed-book tests and two open-book tests. This was instituted to equate participants' encoding and retrieval practice with both tests, as well as to equate the build-up of proactive interference. Second, the three critical test expectancy instructions (open-book, closed-book, and non-specific) were manipulated between participants and a substantial cash award was offered in order to avoid potential motivational differences at retrieval across the three groups.

The main comparison of interest in Experiment 2 was how test expectancy instructions before studying would influence final retention after two days on delayed fact, comprehension, and transfer tests. Results from Experiment 1 suggested that, in the absence of appropriate test expectancy, the type of initial test received does not influence final test performance. In Experiment 2, we hypothesised that the type of final test *expected* would influence delayed test performance to a greater degree than type of initial test *received*. Specifically, we predicted that closed-book test expectancy would encourage effortful studying habits more than open-book test expectancy, resulting in greater delayed fact, comprehension, and transfer test performance.

In addition to the test expectancy manipulation there were three supplemental changes to the procedure used in Experiment 2. First, in order to address the finding that students had greater performance on transfer questions than on comprehension questions in Experiment 1, transfer questions were presented in short answer format (i.e., multiple-choice alternatives were omitted) during Session 2 in Experiment 2. Recall of causal reasons (assessed by the transfer questions) via short answer questions may evoke transfer to a greater degree than multiple-choice questions requiring recognition of causal antecedents. Second, we also included fact test questions during Session 2 in Experiment 2, in order to provide some basis for comparison to the previous study by Agarwal et al. (2008), which used fact-based materials, but which did not include test expectancy as a manipulation. Third, feedback was not presented after initial closed-book tests (in the practice phase), because these tests were used to control for participants' exposure and build-up of proactive interference; providing feedback for participants in the closed-book test condition in order to draw comparisons to the open-book test condition (in which participants receive feedback during the test) was not of primary interest in Experiment 2.

A secondary interest of Experiment 2 was how students would study following non-specific expectancy instructions, and subsequently how students' self-selected encoding/study strategies would influence final retention after two days. At the end of the experiment, participants completed a questionnaire designed to examine the study and retrieval strategies they employed during the initial tests. Participants in the non-specific expectancy group were also asked to report which kind of final test they expected (i.e., studied for), open-book or closed-book. Based on previous literature (Feldhusen, 1961; Theophilides & Koutselini, 2000), we predicted that a majority of students in the non-specific expectancy group would expect a final closed-book test and thus have similar delayed test performance to that of the closed-book test expectancy group.

Finally, regarding the practice passages and tests (completed before the critical manipulation of test expectancy), we expected that initial and delayed test performance would replicate findings from Experiment 1, in that final performance following initial open-book and closed-book practice tests would be similar across delayed test types, even when feedback was not provided after practice closed-book tests.

Method

Participants. A total of 108 subjects ($M_{age} = 20.5$ years old, 70 females) were recruited from the Department of Psychology human subject pool. Participants received cash payment and were informed of a $20 reward for top-scoring participants. Four top-scoring participants received a $20 reward after data collection was completed.

Design. Participants took part in a practice phase followed by a critical phase. For the practice phase a 2 (practice test condition: closed-book, open-book) × 3 (delayed test type: transfer, fact, comprehension) within-participant design was used. The practice phase was similar to the design used in Experiment 1, except that a study-only condition was not included and closed-book tests were not followed by feedback. For the critical phase, a 3 (expectancy: closed-book, open-book, non-specific) × 3 (delayed test type: transfer, fact, comprehension) mixed design was used, such that expectancy was manipulated between participants, whereas delayed test type was manipulated within participants.

Six presentation orders of the passages used in the practice (four passages) and critical (two passages) phases were determined using a Latin Square design. The test conditions during the practice phase followed two orders: (1) closed, open, closed, open, or (2) open, closed, open, closed. A subset of 36 participants were randomly assigned to each of the three between participants critical test expectancy conditions, three participants in each of 12 (6 passage by 2 practice test expectancy) orders.

Materials. The six passages used in Experiment 2 were the same as those used in Experiment 1. For Session 1 all comprehension multiple-choice questions used were identical to those used in Experiment 1. For Session 2 three types of criterial test items were used: transfer, fact, and comprehension. The six transfer test questions per passage were also identical to those used in Experiment 1, except multiple-choice alternatives were not presented (i.e., transfer questions were short answer). In addition, the experimenter developed six short answer fact questions for each passage. For example, a fact question from a passage about William Penn asked, ''In what year did Penn receive a grant of land in America?'' All fact questions were designed to have one- or two-word answers (e.g., a name, location, date, etc.).

Participants completed comprehension tests during the practice phase of the first session of the experiment. During the second session two days later, participants completed closed-book tests in the same order blocked by passage (short answer transfer questions first, short answer fact questions second, and then multiple-choice comprehension questions last). Comprehension questions encountered in Session 2 were identical to those in Session 1 (during the practice phase); however, the five response options were presented in a new random order for the second session. All passages and tests were studied and completed in paper-and-pencil format, and the order of questions for each test was the same across all participants.

Procedure. Participants were tested individually or in small groups. As in Experiment 1, all passages and tests were pre-arranged (according to counterbalancing order) in one blue folder for each participant, face down. An empty red folder was also provided for each participant, for completed passages and tests. Each participant was seated at a computer and used an E-Prime 1.0 program (Schneider et al., 2002) that provided instructions and recorded time spent on each phase of the experiment. The experimenter was outside the testing room and monitored compliance with all instructions via a window.

All study and test periods were self-paced, although a maximum of 4 minutes (ample time, considering the maximum time on task in Experiment 1 was 2.5 minutes) per period during Session 1 was imposed in order to mimic time constraints in typical classroom settings (participants did not receive a time limit during Session 2 tests). Before each study and test period, participants were reminded of the 4-minute time limit, and study and test instructions were identical to those provided in Experiment 1.

In Session 1 participants were instructed that they would read several passages and take multiple-choice tests. They were also instructed that the top-scoring participants would receive a $20 cash reward. Participants first took part in a practice phase, followed by a critical phase, but participants were not informed that the first phase was for practice or that the second phase was ''critical'', in order to maintain similar levels of motivation across all passages and tests.

During the practice phase, expectancy instructions and actual practice tests received always matched (e.g., when a participant expected an open-book test, the passage was followed by an open-book test). Participants read four practice passages, two while expecting a closed-book test, and two while expecting an open-book test. For the closed-book practice test condition (but before reading), participants were instructed, ''After you read this passage, you will receive a closed-book multiple-choice test. While completing the closed-book test, you WILL NOT be allowed to look at the passage. Please keep this in mind while reading the passage at your own pace.'' Participants read the passage and then completed a multiple-choice comprehension test without viewing the passage; feedback was not provided. For the open-book practice test condition (but before reading), participants were instructed, ''After you read this passage, you will receive an open-book multiple-choice test. While completing the open-book test, you WILL be allowed to look at the passage. Please keep this in mind while reading the passage at your own pace.'' Participants read the passage and then were allowed to view the passage while completing a multiple-choice comprehension test.

Following the practice phase, the critical phase of Session 1 began in which different sets of participants received either closed-book, open-book, or non-specific instructions pertaining to Session 2. Participants in the closed-book test expectancy group were instructed, ''Before you read the next passage, it is important to mention that you will receive a closed-book test on this passage during your NEXT session in two days. While completing the closed-book test, you WILL NOT be allowed to look at the passage during your next session. Please keep this in mind while reading the passage at your own pace.'' Participants in the open-book test expectancy group were instructed, ''Before you read the next passage, it is important to mention that you will receive an open-book test on this passage during your NEXT session in two days. While completing the open-book test, you WILL be allowed to look at the passage during your next session. Please keep this in mind while reading the passage at your own pace.'' Participants in the non-specific expectancy group were instructed, ''Before you read the next passage, it is important to mention that you will receive a test on this passage during your NEXT session in two days. Please keep this in mind while reading the passage at your own pace.'' After the instructions were presented according to expectancy group, participants read one passage. The expectancy instructions were repeated and then participants read a second passage under the same conditions. After reading the second critical passage, participants were reminded to return two days later for Session 2 and were dismissed (i.e., participants did not complete initial tests on the critical passages after receiving expectancy instructions and reading).

Session 2 occurred two days after Session 1. Participants completed short answer transfer, short answer fact, and multiple-choice comprehension tests over all six passages without restudying or reviewing the passages (i.e., the final tests were closed-book) at their own pace (no time limit was imposed). Tests were blocked by passage, such that when a participant finished the transfer, fact, and comprehension tests for one passage (in that order), they moved on to the transfer, fact, and comprehension tests for the next passage. Tests on critical passages were followed by tests for the remaining four practice passages (in order to measure performance following initial tests, to confirm results from Experiment 1), in the order in which participants encountered the critical and practice passages during Session 1.

Finally, participants were asked to complete a short questionnaire about study/test strategies and prior experience with closed-book and open-book tests (see Appendix; question 4 adapted from Farr, Pritchard, & Smitten, 1990). Participants in the non-specific expectancy group were asked, at the end of the experiment, which kind of test they actually expected during Session 2; due to experimenter error, however, these data were not collected from 8 of the 36 of participants in the non-specific expectancy group. The entire procedure lasted approximately 120 minutes across the two sessions. All participants were debriefed and thanked for their time.

Results

Initial test performance. Initial closed-book and open-book practice test performance is shown in the first column of Table 3. Replicating Experiment 1, initial test performance was significantly better on open-book tests ($M = .68$) in comparison to closed-book tests ($M = .62$), $F(1, 105) = 7.76$, $\eta_p^2 = 0.07$.

TABLE 3
Initial and delayed test performance (proportion correct) for practice test conditions in Experiment 2

	Initial Comprehension Test	*Delayed Fact Test*	*Delayed Comprehension Test*	*Delayed Transfer Test*
Practice closed-book tests	.62 (.02)	.22 (.01)	.63 (.01)	.40 (.02)
Practice open-book tests	.68 (.02)	.20 (.01)	.66 (.02)	.42 (.02)

Delayed tests occurred two days after initial comprehension tests. Standard errors of the mean are displayed in parentheses, $n = 108$.

Delayed test performance. Final test performance for the practice closed-book and open-book test conditions is shown in Table 3. Consistent with Experiment 1, delayed multiple-choice comprehension test performance was similar for passages that were tested immediately in an open-book format ($M = .66$) and in a closed-book format ($M = .63$). Similar patterns held for the delayed short answer fact test ($M = .20$ for passages initially tested in an open-book format, and $M = .22$ for passages tested in a closed-book format) and the delayed short answer transfer test ($M = .42$ for passages initially tested in an open-book format, and $M = .40$ for passages tested in a closed-book format). Three separate one-way ANOVAs on practice test condition (open-book, closed-book) for each of the three delayed test types (transfer, fact, comprehension) did not reveal any differences between the two practice test conditions, Fs < 2.48, ps $> .118$. The present results confirm our hypothesis that type of initial test received, open-book or closed-book, does not influence final retention measured two days later, even when participants received feedback (via the passage) in the open-book practice condition, but did not receive feedback following the closed-book practice condition.

Before turning to the key comparisons in delayed recall among the three critical conditions (open-book, closed-book or non-specific expectancy conditions) that were implemented in Session 1, we turn to one other matter: what kind of tests participants reported expecting in the non-specific expectancy condition. Consistent with our a priori hypotheses, 25 of the 36 participants in the non-specific expectancy group reported that they expected a final closed-book test, 3 reported expecting an open-book test, and perceived expectancy from 8 participants was not collected due to experimenter error. Delayed test performance for the non-specific, closed-book, and open-book expectancy groups is reported in Table 4. The data show little, if any, difference between performance in the non-specific and closed-book expectancy groups, but both groups have consistently greater performance than the open-book expectancy group. Indeed, three univariate ANOVAs on delayed comprehension, transfer, and fact test performance for the non-specific and closed-book expectancy conditions revealed no differential effects of expectancy on delayed performance, Fs < 1, ps $> .389$. The distribution of responses (i.e., 89% of participants from whom data was collected expected a closed-book final test in the non-specific expectancy group) and similar levels of performance between the non-specific and closed-book expectancy groups confirmed our hypothesis that, in the absence of specific test instructions, a majority of students expected and subsequently studied for closed-book tests.

Our initial predictions confirmed, we continued with planned comparisons between a group comprising both the non-specific and closed-book expectancy participants ($n = 72$, hereafter referred to as the "combined closed-book expectancy group") and the open-book expectancy group ($n = 36$).[1] Overall, delayed test performance for the combined closed-book expectancy group ($M = .39$) was greater than delayed test performance for the open-book expectancy group ($M = .33$), $F(1, 97) = 14.02$, $\eta_p^2 = 0.13$, suggesting that participants preparing for a closed-book test seem to study harder (or more effectively) than those expecting an open-book test.

Three separate univariate ANOVAs for critical expectancy condition (combined closed-book,

[1]Separate analyses between the non-specific and open-book, and also the closed-book and open-book expectancy groups provided either marginally significant or non-significant results due to relatively low power. The results in Tables 4 and 5 show that the results of the closed-book and non-specific expectancy condition are quite similar, so we combined data from these two groups for greater statistical power. Because a majority of participants in the non-specific expectancy group reported expecting a closed-book test, this practice seems justified.

TABLE 4
Delayed test performance (proportion correct) for test expectancy conditions in Experiment 2

	Delayed Fact Test	*Delayed Comprehension Test*	*Delayed Transfer Test*	*Delayed Average*
Non-specific expectancy group ($n = 36$)	.23 (.03)	.60 (.03)	.38 (.03)	.40 (.02)
Closed-book expectancy group ($n = 36$)	.21 (.03)	.57 (.03)	.38 (.04)	.39 (.02)
Open-book expectancy group ($n = 36$)	.18 (.02)	.53 (.02)	.29 (.03)	.33 (.02)

Delayed tests occurred two days after initial comprehension tests. Standard errors of the mean are displayed in parentheses.

open-book), one for each delayed test type (transfer, fact, comprehension) were conducted. On the delayed fact test the difference between the combined closed-book expectancy condition ($M = .22$) and the open-book expectancy condition ($M = .18$) was not significant, $F(1, 107) = 1.98$, $p = .162$. For the delayed comprehension test the difference between the combined closed-book expectancy group ($M = .59$) and the open-book expectancy group ($M = .53$) was marginally significant, $F(1, 107) = 3.07$, $p = .083$, $\eta_p^2 = 0.03$. Finally, the combined closed-book expectancy group ($M = .38$) showed almost a 10% increase on delayed transfer test performance over the open-book expectancy group ($M = .29$), $F(1, 107) = 5.31$, $\eta_p^2 = 0.05$, demonstrating the large benefit from closed-book test expectancy on a delayed transfer test.

Time on task. Time on task for the two within-participant practice conditions (closed-book and open-book) and the three between-participants expectancy groups (non-specific, closed-book, and open-book) is shown in Table 5. Due to computer error during Session 1, response times for a few passages and tests were not recorded; thus the number of participants included in each analysis varies. During the first session there was no significant difference in time spent reading passages for the practice open-book ($M = 130$ s) and closed-book ($M = 136$ s) test conditions, $F(1, 102) = 1.56$, $p > .05$ ($n = 103$), which is not surprising because participants did not know which kind of immediate test to expect before or during the study phase. The time spent completing the practice open-book comprehension tests ($M = 168$ s) was significantly greater than time spent completing the practice closed-book

TABLE 5
Time on task (seconds) in Experiment 2

	Session 1		*Session 2*	
	Reading	*Initial Comprehension Test*	*Delayed Comprehension Test*	*Delayed Transfer and Fact Test*
Practice closed-book tests ($n = 108$)	136.4 (104)	148.2 (107)	109.6	265.7
Practice open-book tests ($n = 108$)	129.7 (106)	168.1 (81)	106.9	270.2
Non-specific expectancy group ($n = 36$)	163.6 (31)		168.8	398.0
Closed-book expectancy group ($n = 36$)	165.7 (32)		146.5	397.5
Open-book expectancy group ($n = 36$)	127.7 (35)		173.1	370.9

Session 2 occurred two days after Session 1. Due to computer error, response times for some passages and tests during Session 1 were not recorded; the number of participants in each group is displayed in the left-hand column, and the number of participants in each average is displayed in parentheses.

comprehension tests ($M = 148$ s), $F(1, 79) = 5.49$, $\eta_p^2 = 0.07$ ($n = 80$), suggesting that participants made use of the available passage while completing the open-book practice tests.

Turning to the two critical between-participants expectancy groups (again, data were combined for participants in the non-specific and closed-book expectancy groups, but see Table 5 for means for all three expectancy groups), a between-participants effect of test expectancy on reading times for the critical passages during Session 1 was demonstrated, $F(1, 96) = 14.01$, $\eta_p^2 = 0.13$ ($n = 98$). Participants tailored their effort based on the final test expectancy instructions provided and spent significantly less time reading the passage when expecting a final open-book test ($M = 128$ sec) than when expecting a final closed-book test ($M = 165$ s). This outcome provides direct evidence that participants study less hard (or at least for less time) when expecting an open-book test.

Response times for Session 2 for all participants are also shown on the right side of Table 5 (no computer errors occurred when collecting Session 2 data). The E-Prime program used in the present experiment collected total time spent on the short answer test, thus time spent on the individual transfer and fact tests are collapsed. Two separate univariate ANOVAs on practice test condition (open-book, closed-book) for each delayed test type (short answer transfer/fact, multiple-choice comprehension) did not reveal any differences between time spent completing delayed tests across the two practice conditions, Fs < 1, ps $> .686$. Similarly, separate univariate ANOVAs on expectancy conditions (combined closed-book, open-book) for each delayed test type did not reveal any differences between time spent completing delayed tests for the two expectancy conditions, Fs < 1.21, ps $> .274$. Thus, for Session 2, type of initial practice test condition and test expectancy instructions did not influence the amount of time spent completing the final tests; the only difference was that participants took longer to complete delayed short answer tests ($M = 308$ s) than multiple-choice tests ($M = 190$ s), which is hardly a surprise.

Questionnaire results. Of the 108 participants who took part in Experiment 2, 51 (47%) preferred open-book class examinations; 33 (31%) preferred closed-book exams; and 24 (22%) had no preference, but three univariate ANOVAs revealed no interaction between test preference and performance on the three delayed test types (on all passages, practice and critical), Fs < 1.86, ps $> .05$. The two most common reasons participants provided for preferring open-book exams was that they felt they were easier than closed-book exams ($n = 16$) and that they favoured using reference material during an exam ($n = 16$). The most commonly cited reason for preferring a closed-book exam ($n = 13$) was that participants felt that closed-book exam material is typically easier than open-book exam material.

The most commonly reported strategy ($n = 50$) used during open-book tests was reading the entire passage, reading/answering each question, and then checking the passage for correct answers. A total of 27 participants reported partially reading the passage before answering questions, 21 participants reported reading all questions before reading the entire passage, 6 participants reported alternating between reading/answering questions and searching the passage for answers, and 4 participants reported other strategies. Similar to test preference, strategy use did not interact with performance on the three delayed test types (on all passages), Fs < 1.73, ps $> .05$.

Discussion

Consistent with Experiment 1 and Agarwal et al. (2008), initial open-book and closed-book tests did not produce different levels of delayed test performance. Importantly, participants adjusted their study time during Session 1 in accordance with final test expectations, which resulted in a similar pattern of delayed test performance during Session 2: participants spent the least time reading passages and had the lowest delayed transfer, fact, and comprehension test performance when provided open-book test expectancy instructions. In other words, closed-book test expectancy increased delayed transfer test performance by almost 10% in comparison to open-book test expectancy. We also note that overall delayed test performance following initial tests ($M = .42$) was greater than delayed test performance following test expectancy instructions ($M = .37$), $t(107) = 3.90$, $d = 0.41$, confirming the benefits of initial testing on long-term learning, over and above the influence of test expectancy.

GENERAL DISCUSSION

The current experiments provide additional insight into the effects of completing open-book and closed-book tests for enhancing long-term learning. In Experiment 1 we obtained significant testing effects following both open-book tests and closed-book tests with feedback (in comparison to the study-only condition), although the two types of initial tests did not produce differential benefits for long-term retention or transfer of knowledge (replicating Agarwal et al., 2008), even when using materials specifically designed for open-book tests (e.g., GRE passages and tests). In Experiment 2 the critical role of test expectancy regarding open-book and closed-book tests was examined. The main finding was that participants studied for less time when expecting an open-book test and correspondingly performed worse on later retention tests. Expecting an open-book test (relative to a closed-book test) led to worse performance on all three types of items on the delayed criterial test (transfer, fact, and comprehension questions), with a 9% decrease in performance on transfer items, a marginal 6% decrease in performance on comprehension items, and a non-significant 4% decrease in performance on fact items.

A topic of interest throughout the current project was whether open-book versus closed-book tests would differentially benefit participants' ability to transfer knowledge from initial comprehension questions to final transfer questions that required causal reasoning, using GRE materials designed for open-book testing. In Experiment 1 both types of initial test conditions increased transfer performance by 10% relative to the study-only control condition. In Experiment 2, however, closed-book *expectancy* increased delayed transfer test performance by almost 10% relative to open-book test expectancy, although it remains unclear why a similar benefit was not demonstrated for delayed comprehension and fact test performance. In other words, students' expectations of a final closed-book test enhanced delayed transfer performance, even when an initial test was not completed. This demonstrates the robust effect of test expectancy on later performance.

Some education researchers (e.g., Jacobs & Chase, 1992) have argued that open-book tests promote transfer relative to closed-book tests; our findings are inconsistent with this idea. Our results seem more consistent with the constructivist theory that if a reader is not required to construct a meaningful situation model, transfer will not occur. In the case when students expect open-book tests and can rely on the passage for information, he or she may not generate inferences (Graesser et al., 1994). Because inferences were a necessary component of our transfer questions, the relative failure to generate situation models may have caused poorer retention. However, this interpretation is hypothetical at this point and in need of future research (because we did not measure development of situation models following the two types of tests). Still, our results show no benefit of initial open-book relative to closed-book tests on any of the three types of questions used on a delayed (closed-book) criterial test.

A second issue of interest in the current research was the potential influence of test expectancy on students' study behaviours and delayed test performance. In real-world educational settings, students are often informed about the type of questions and exams to expect during a course (e.g., multiple-choice, short answer, essay, closed-book, open-book, take home, etc.), but previous work on open-book and closed-book tests (Agarwal et al., 2008) did not investigate this critical component of testing. In Experiment 2 we hypothesised that (1) in the absence of test expectancy, students mostly study for closed-book tests, and (2) in the presence of test expectancy, students expend more time and/or effort while studying for a closed-book test than for an open-book test.

Confirming the first hypothesis, a majority of participants in the non-specific expectancy group later reported actually expecting a closed-book test and delayed test performance was similar for the non-specific and closed-book expectancy groups. Confirming the second hypothesis, participants in the open-book expectancy group spent the least amount of time reading during the first session and had the lowest level of delayed performance, particularly for transfer tests, in the second session. We conclude that non-specific and closed-book test expectancy instructions increase students' study time and subsequently enhance delayed test performance.

Although initial retrieval practice during a closed-book test has no different effect from practice on an open-book test, the *expectation* of a final closed-book test seems to be a more potent factor influencing long-term learning. Perhaps

students' prior experience with both open-book and closed-book tests leads them to judge closed-book tests as more difficult, and students subsequently adapt their study habits accordingly in preparation for a later test by spending more time studying. From the present experiments, however, it remains unclear how participants employed the extra study time following non-specific or closed-book test expectancy instructions. Although students typically report that they reread notes or textbook chapters while studying (e.g., Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007), additional research should be conducted to examine what, specifically, students do *while* reading and studying.

The current work adds to a growing literature regarding how to improve learning through the appropriate use of student-driven or teacher-driven strategies (Rohrer & Pashler, 2010; Willingham, 2009). In addition to the use of open-book tests in educational settings, other pedagogical strategies common to education are often ineffective or at least less effective than control comparisons. These techniques include the overlearning of maths procedures (Rohrer & Taylor, 2006), the withholding of feedback (Butler et al., 2008; Marsh, Agarwal, & Roediger, 2009), concept mapping (Karpicke & Blunt, 2011), and the blocking of practice problems (Kornell & Bjork, 2008; Taylor & Rohrer, 2010).

In sum, the present experiments demonstrate that test expectancy can influence study and test behaviours more than type of initial open-book or closed-book retrieval practice. Based on our results we recommend that teachers administer frequent quizzes, because testing improved long-term retention in both experiments reported here (and in many others; see Roediger et al., 2010). Of course we also recommend that teachers give closed-book tests or at least do not announce in advance that they will be giving open-book tests. Simply put, students' study habits may be based, in large part, on the perceived difficulty of a final test; that is open-book versus closed-book test expectancy instructions appear to drive differences in delayed retention and transfer more than open-book versus closed-book initial retrieval practice. Even if teachers plan to give an open-book test, they would be better off not to tell students this fact and provide no specific test expectancy (or otherwise students will not study very much or very effectively for the test).

Manuscript received 25 August 2010
Manuscript accepted 18 July 2011
First published online 14 October 2011

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.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *36*, 1118–1133.

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 918–928.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory and Cognition*, *34*, 268–276.

College Board. (2009). *The official SAT study guide* (2nd ed.). New York, NY: College Board.

Duchastel, P. C., & Nungester, R. J. (1982). Testing effects measured with alternate test forms. *Journal of Educational Research*, *75*, 309–313.

Educational Testing Service. (2002). *Preparing for the verbal and quantitative sections of the GRE general test: Sample questions with explanations.* Princeton, NJ: Educational Testing Service.

Eilertsen, T. V., & Valdermo, O. (2000). Open-book assessment: A contribution to improved learning? *Studies in Educational Evaluation*, *26*, 91–103.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, *27*, 209–226.

Feldhusen, J. (1961). An evaluation of college students' reactions to open book examinations. *Educational and Psychological Measurement*, *21*, 637–646.

Feller, M. (1994). Open-book testing and education for the future. *Studies in Educational Evaluation*, *20*, 235–238.

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

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.

Ioannidou, M. K. (1997). Testing and life-long learning: Open-book and closed-book examination in a university course. *Studies in Educational Evaluation*, *3*, 131–139.

Jacobs, L. C., & Chase, C. I. (1992). *Developing and using tests effectively: A guide for faculty.* San Francisco, CA: Jossey-Bass Publishers.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (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., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772–775.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*, 471–479.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219–224.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the ''enemy of induction? *Psychological Science*, *19*, 585–592.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*, 212–218.

Marsh, E. J., Agarwal, P. K., & Roediger, H. L. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied*, *15*, 1–11.

Neely, J. H., & Balota, D. A. (1982). *On inferring test-expectancy-induced encoding differences in recall and recognition: A methodological, empirical, and theoretical analysis.* Unpublished manuscript, Purdue University, West Lafayette, IN.

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35*, 1917–1927.

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*, 559–579.

Research & Education Association, Inc. (2008). *The very best coaching and study course for the GRE General Test*. Piscataway, NJ: Research & Education Association, Inc.

Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New frontiers in applied memory* (pp. 13–49). Hove, UK: Psychology Press.

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

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher*, *39*, 406–412.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209–1224.

Schmidt, S. R. (1980). *Test expectancy and test appropriate processing in the retention of prose.* Unpublished doctoral dissertation, Washington University, St. Louis, MO.

Schmidt, S. R. (1983). The effects of recall and recognition test expectancies on the retention of prose. *Memory & Cognition*, *11*, 172–180.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837–848.

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and open-book examination: A comparative analysis. *Educational Research and Evaluation*, *6*, 379–393.

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*, 240–245.

Willingham, D. T. (2009). *Why don't students like school: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco, CA: Jossey-Bass Publishers.

APPENDIX

Experiment 2 Questionnaire

(1) Have you taken a closed-book test before this experiment? (circle one) Yes No

If so, when was the last time you took a closed-book test for a class?

_____

(2) Have you taken an open-book test before this experiment? (circle one) Yes No

If so, when was the last time you took an open-book test for a class?

_____

(3) Which do you prefer for class examinations? (circle one)

Closed-book test Open-book test No preference

_____

Why? _____

(4) What do you *primarily* do during an open-book test? (check only one)

☐ Read passage, then read each question, then search the passage for the correct response

☐ Partially read passage, then read each question, then search the passage for the correct response

☐ Read all questions, then read the entire passage, then reread each question and search the passage for the correct response

☐ Read the first question, then search the passage for the correct response, then move on to the next question

☐ Other (please explain): _____

_____