January 2009

Test Expectancy and Transfer of Knowledge with Open-Book and Closed-Book Tests

Pooja Agarwal
Washington University in St. Louis

WASHINGTON UNIVERSITY

Department of Psychology

TEST EXPECTANCY AND TRANSFER OF KNOWLEDGE

WITH OPEN-BOOK AND CLOSED-BOOK TESTS

by

Pooja Kay Agarwal

A thesis presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the
degree of Master of Arts

December 2009

Saint Louis, Missouri

Acknowledgements

I wish to thank the members of my Thesis Examination Committee, Dr. Henry L. Roediger, III, Dr. David A. Balota, and Dr. Kathleen B. McDermott, for their guidance and support throughout this project. I also thank Jake Sanches, Emily Rosenzweig, and Jane McConnell for their assistance, as well as Andrew Butler for valuable discussions. This research was supported by the National Science Foundation Graduate Research Fellowship Program 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.

Table of Contents

List of Tables

Abstract

Two experiments examined the testing effect with open-book tests, in which students view notes and textbooks while taking the test, and closed-book tests, in which students take the test without viewing notes or textbooks. In the first experiment, subjects studied GRE passages and then took an open- or closed-book test. Open-book testing led to better initial performance than closed-book testing, but this benefit did not persist and both types of testing produced equivalent retention on GRE comprehension questions and transfer questions after a two-day delay. In the second experiment, subjects were informed in advance of the type of immediate or delayed test to expect, in order to mimic educational settings in which students typically know the type of quiz or exam to expect with regards to open-book vs. closed-book format. Initial retrieval practice during these two types of tests did not yield differences in long-term retention (consistent with Experiment 1), but final test expectancy significantly influenced delayed retention. Closed-book test expectancy produced greater final test performance on comprehension, transfer, and factual questions in comparison to open-book test expectancy, demonstrating that test expectancy can influence long-term learning.

Test Expectancy and Transfer of Knowledge with Open-Book and Closed-Book Tests

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 our memory for information (e.g., Roediger & Karpicke, 2006a). This *testing effect* has been well established in the literature and recent research has focused on educational implications of the testing effect (for reviews, see Roediger & Karpicke, 2006b; Roediger, Agarwal, Kang, & Marsh, in press).

For instance, Kang, McDermott, and Roediger (2007, Experiment 2) evaluated the testing effect with short answer and multiple-choice tests. During the initial session, subjects studied four papers from *Current Directions in Psychological Science*, and then either completed an immediate multiple-choice test (followed by correct answer feedback), completed an immediate short answer test (followed by correct answer feedback), read a list of statements (corresponding to the test answers), or completed a filler task. After three days, subjects returned for a second session and received a test with both multiple-choice and short answer questions for each paper. Significant testing effects were found, such that taking an initial multiple-choice or short answer test enhanced final performance in comparison to the control (filler) condition. Specifically, an initial short answer test produced the greatest final test performance on both multiple-choice and short answer final test questions. Kang et al. concluded that short answer quizzes (followed by feedback) are more effective than multiple-choice quizzes in enhancing student learning.

There exists another distinction between quiz formats commonly found in educational settings: closed-book and open-book quizzes. During a closed-book quiz or test, students are not allowed to refer to notes or textbook materials. During an open-book quiz, however, students are allowed to refer to notes or textbook materials. Although these two types of quizzes are usually used for assessment purposes, we can ask a similar question as Kang et al. asked: which quiz 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 subjects read a series of passages, which were followed by a closed-book test, a closed-book test with feedback (where subjects graded their own responses), or an open-book test. Subjects also read a passage in the study-only condition, which was not followed by a test. After one week, subjects returned for a second session and completed closed-book tests over each passage studied during the first session. Across two experiments (average data displayed in Table 1), 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 final performance after one week, and final performance following these tests was greater than performance following the study-only condition, i.e., a significant testing effect.

One possible criticism of Agarwal et al. is that the materials used were not appropriate for open-book tests, since open-book tests are supposed to enable a student to integrate and transfer information (Jacobs & Chase, 1992). Materials that are fact-based, such as the prose passages and short answer tests used by Agarwal et al., may not reveal a benefit following open-book testing. Instead, instructors maintain that open-book tests are

Table 1

Initial and final test performance (proportion correct) averaged across Experiments 1 and 2 of Agarwal et al. (2008)

	Initial Test	Delayed Test
Study-only		.43
Closed-book test	.70	.57
Closed-book test with feedback	.67	.67
Open-book test	.81	.66

Note. The delayed test occurred one week after initial tests.

designed to require students to apply knowledge, rather than memorize or restate it; therefore, if a student uses higher-order thinking skills during an initial open-book test, benefits for final retention may emerge (Feller, 1994; Theophilides & Koutselini, 2000).

On the other hand, the transfer appropriate processing framework and the concept of desirable difficulty indicate that an initial closed-book test (followed by feedback) should produce greater final closed-book test performance than an initial open-book test, regardless of type of material. First, transfer appropriate processing suggests that when processes engaged during encoding match processes required at retrieval, memory is enhanced (Bransford, Franks, Morris, & Stein, 1979). Thus, an initial closed-book test should result in better performance on a final closed-book test, in comparison to an initial open-book test. In addition, the concept of desirable difficulty suggests that more challenging test conditions may slow initial learning, but ultimately result in enhanced final performance (Bjork, 1994). In this case, 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 frequently spend less time studying for open-book tests (Eilertsen & Valdermo, 2000; Jacobs & Chase, 1992).

Thus, any differences in memorial benefits between open-book and closed-book tests have yet to be determined. In order to further our understanding of the testing effect with open-book and closed-book tests, the current two experiments include materials designed for open-book testing: comprehension and transfer questions. Specifically, initial passages and tests were drawn from a Graduate Record Examination (GRE) test preparation book. Comprehension questions from the verbal section of the GRE require students to analyze relationships, apply the author's ideas to novel situations, and draw

inferences (Educational Testing Service, 2002). During GRE comprehension questions, students are both allowed and required to refer back to the passage at hand, akin to an open-book test.

Final tests in the current experiments were comprised of both original GRE comprehension questions, as well as higher-order transfer questions where subjects answered "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 casual 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 colonization of Pennsylvania (see Appendix A for an example passage, comprehension test, and transfer test on William Penn) included the following:

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

A corresponding final transfer 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

In order to draw distinctions between these two types of questions, we used Barnett and Ceci's (2002) Taxonomy of Transfer. They included two main factors or areas in which transfer can occur: content (what is transferred) and context (when and where transfer occurs). In terms of the content factor, Barnett and Ceci included transfer of a learned skill (e.g., a problem-solving heuristic or procedure), performance change in speed or accuracy, and transfer of memory demands (e.g., recognizing to recalling). In terms of the context factor, Barnett and Ceci included six dimensions: knowledge, physical context, temporal context, functional context, social context, and modality. For instance, transfer in the knowledge domain would include going from biology topics to economics; physical transfer includes going from school to home; a temporal context would be transfer from one day to the next; and so on.

The types of comprehension and transfer questions used in these experiments fall under the "memory demand" content factor and the "knowledge domain" context factor. Regarding the transfer of a memory demand, subjects initially answered comprehension questions based on specific ideas from the passage and then answered transfer questions that required causal understanding and reasoning. Subjects were not explicitly informed that transfer questions were related to comprehension questions, but because subjects received transfer tests following comprehension tests, subjects probably recognized the association and recalled related information. As Barnett and Ceci explained, this

procedure (i.e., without hints about related information) requires recall of a learned skill and its applicability, as well as the ability to execute the required memory demand and transfer it to a new task.

Regarding transfer of context within the knowledge domain, Barnett and Ceci (2002) described the knowledge domain as "the knowledge base to which the skill is to be applied" (p. 623). For example, transfer of knowledge about mice to rats was classified as "near" transfer by Barnett and Ceci, whereas transfer from a science class to an art class was classified as "far" transfer within the knowledge domain. Because the comprehension and transfer questions in this 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 probably more salient than the transfer of knowledge between the comprehension and transfer questions.

In contrast to a specific taxonomy of transfer categories, Bloom's Taxonomy of Educational Objectives included six categories of cognitive domains, Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation, ordered from simple and concrete to complex and abstract, and it is currently used to classify assessment items (Krathwohl, 2002). The comprehension questions in the current study fit within the Comprehension category, whereas the transfer questions fit within the Analysis category. Krathwohl's Revised Taxonomy (2002) includes two dimensions: Knowledge and Cognitive Process. Within the Knowledge dimension, our comprehension questions are factual, whereas our transfer questions are conceptual. Within the Cognitive Process dimension, our comprehension questions require Understanding, whereas the transfer

questions require Analyzing processes. By deliberately using two different types of questions, detailed GRE comprehension questions and conceptual/inferential transfer questions, we hoped to evaluate any potential benefits for open-book testing that did not emerge with Agarwal et al.'s (2008) fact-based materials.

In Experiment 1, we aimed to extend Agarwal et al. by replicating results for three key conditions (study-only, closed-book test with feedback, and open-book test) while measuring performance on both comprehension and transfer questions. Performance on final comprehension and transfer questions may be greater following an initial open-book test than an initial closed-book test, because open-book testing may promote high-order cognitive skills, consistent with the educational literature. Alternatively, performance on final comprehension and transfer questions may be greater following an initial closed-book test, because closed-book testing may require more challenging processing, consistent with the transfer appropriate processing framework and the concept of desirable difficulty.

Experiment 1

Method

Participants. Seventy-two subjects were recruited from the Washington University in St. Louis Department of Psychology human subject pool. Subjects received either credit towards completion of a research participation requirement or cash payment.

Design. The three within-subjects initial learning conditions (study-only, closed-book test with feedback, open-book test) crossed with two dependent variables (final comprehension test questions, final transfer test questions) are displayed in Table 2. Six passages were presented in the same order for all subjects, but the order in which the

Table 2

Learning conditions in Experiment 1

	Session 1	Session 2
Study-only	Study	Comprehension Test
Closed-book test with feedback	Study, Test, Self-grade test (with passage available)	Comprehension Test
Open-book test	Study, Test (with passage available)	Comprehension Test
Study-only	Study	Transfer Test
Closed-book test with feedback	Study, Test, Self-grade test (with passage available)	Transfer Test
Open-book test	Study, Test (with passage available)	Transfer Test

Note. All tests during Session 1 were comprehension tests. Session 2 occurred two days after Session 1. Subjects completed final closed-book tests over each passage in Session 2.

conditions occurred was counterbalanced using a Latin Square. All conditions appeared once in every ordinal position, creating six counterbalancing orders, and twelve subjects were randomly assigned to each of the six orders. Once subjects completed one condition (e.g., studying and taking an open-book test), they moved on to the next condition, according to their counterbalancing order.

Materials. Six passages, approximately 425 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 tests were adapted from the same test preparation book, whereas the experimenter created final multiple-choice transfer tests such that subjects were asked to choose "why" a particular idea that was stated in the passage and asked on the initial comprehension test was true (see Appendix A for an example passage, comprehension test, and transfer test on William Penn). All questions on comprehension tests had a corresponding "why" question on the transfer tests. In addition, all tests were composed of six multiple-choice questions with five-alternative forced choice responses.

Subjects completed comprehension tests during the first session of the experiment. During the second session two days later, subjects completed both comprehension and transfer tests. Comprehension questions encountered in Session 2 were identical to those in Session 1; however, the order of the five alternative forced choice responses was randomly reordered for the second session. All passages and tests were presented and completed in paper-and-pencil format.

Procedure. Subjects were tested individually or in small groups. In Session 1, they were instructed that they would read several prose passages, which might or might not be followed by a test. Thus, subjects did not know whether to anticipate a closed-book test, an open-book test, or another passage. While all passages and tests were presented in a paper-and-pencil format, subjects were seated at a computer and used an E-Prime 1.0 program (Schneider, Eschman, & Zuccoloto, 2002), which provided instructions and recorded time spent on each phase of the experiment.

All study and test periods were self-paced. During a study period, the computer instructed subjects to take a passage from a blue folder, read it at their own pace, place it face down in a 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, subjects were asked to take the corresponding test from the same blue folder (passages and tests were pre-arranged according to counterbalance order for each subject), circle a multiple-choice alternative for every question, place the test face down in the red folder when they completed the test, and push spacebar for the next set of instructions (recording time during testing). The experimenter observed compliance with all instructions provided.

During Session 1, subjects 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. In the study-only condition, subjects read the passage one time and were not tested on it; the computer instructed subjects to move on to the next passage. In the closed-book test with feedback condition, subjects read the passage, completed the multiple-choice comprehension test without viewing the passage, and then subjects were asked to take the

corresponding passage out of the red folder and check their answers. Specifically, subjects were instructed to write "correct" next to responses they believed were correct (based on information from the passage; subjects 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, subjects 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. In Session 2, subjects completed multiple-choice tests over all six passages without restudying or reviewing the passages (i.e., the final tests were closed-book); three tests were repeated comprehension tests and three tests were transfer tests (one for each initial learning condition). At the end of the experiment, subjects were debriefed and thanked for their time.

Results

Initial Test Performance. Initial test performance is show in Table 3. As expected, initial test performance was significantly greater on open-book tests (M = .69) in comparison to closed-book tests (M = .60 feedback was provided after performance was measured), $F(1, 71) = 13.34$, $\eta_p^2 = .16$. For the closed-book test with feedback condition, subjects accurately self-graded 77% of items. Specifically, subjects 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 subjects), indicating that subjects were processing feedback during the self-grading process (even if not perfectly).

Table 3

Initial and final test performance (proportion correct) in Experiment 1

	Initial Comprehension Test	Delayed Comprehension Test	Delayed Transfer Test	Delayed Average (by condition)
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 (by test)	.65 (.02)	.58 (.02)	.67 (.02)	

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

Final Test Performance. Final test performance is shown in Table 3. A 2 (test type: comprehension, transfer) by 3 (learning condition: study-only, closed-book, open-book) analysis of variance (ANOVA) revealed a significant effect of test type, $F(1, 71) = 15.63$, $\eta_p^2 = .18$, and a significant effect of learning condition, $F(2, 142) = 15.36$, $\eta_p^2 = .18$. In general, delayed transfer test performance ($M = .67$) was greater than delayed comprehension test performance ($M = .58$), and the closed-book and open-book learning conditions ($M = .66$ and $M = .67$, respectively) resulted in greater delayed performance than the study-only condition ($M = .55$), $ps < .05$.

After two days, comprehension test performance 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, confirmed by a significant main effect of learning condition on delayed comprehension test performance, $F(2, 142) = 9.24$, $\eta_p^2 = .12$. Comprehension performance for the open-book test condition was significantly greater than performance for the study-only condition, $t(71) = 3.80$, $d = .58$, and comprehension performance for the closed-book test with feedback condition was also greater than the study-only condition, $t(71) = 3.98$, $d = .54$. These results demonstrate the memorial benefit for testing compared to studying, 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, $p > .05$.

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 ($M = .60$), confirmed by a significant effect of learning condition on final transfer test performance, $F(2, 142) = 5.70$, $\eta_p^2 = .07$. Again,

performance for open-book test condition was significantly greater than performance for the study-only condition, $t(71) = 2.85$, $d = .42$, and performance for the closed-book test with feedback condition was also greater than performance for the study-only condition, $t(71) = 2.79$, $d = .39$, confirming the robust effects of testing on transfer of knowledge.

Response Times. Response times are shown in Table 4. During the first session, there were no significant differences in reading times across the study-only ($M = 144.2$ sec), open-book ($M = 145.2$ sec), or closed-book ($M = 148.1$ sec) learning conditions, as would be expected because subjects did not know whether to expect a test before reading the passage, $p > .05$. Time spent completing comprehension tests was greater when subjects were taking open-book tests ($M = 254.0$ sec) in comparison to closed-book tests ($M = 157.2$ sec), $F(1, 71) = 106.99$, $\eta_p^2 = .60$, suggesting that subjects made use of the available passage while completing the open-book test. Subjects 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 ($M = 330.9$ sec) was significantly greater than time spent testing and processing feedback in the open-book condition ($M = 254.0$ sec), $F(1, 71) = 65.78$, $\eta_p^2 = .48$.

For the second session, a 2 (test type: comprehension, transfer) by 3 (learning condition: study-only, closed-book, open-book) ANOVA revealed a significant interaction between test type and learning condition on response time, $F(2, 142) = 17.33$, $\eta_p^2 = .20$. As can be seen from Table 4, time spent on the delayed comprehension test was greater than time spent on the transfer test for the study-only condition, whereas time spent on the transfer test was greater than time spent on the comprehension test for the

Table 4

Response times (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

Note. Session 2 occurred two days after Session 1.

closed-book with feedback condition. Time spent on both types of tests, however, was nearly equivalent for the open-book test condition.

A main effect of learning condition on time spent completing the final comprehension tests was confirmed, $F(2, 142) = 24.70$, $\eta_p^2 = .26$, such that time spent was significantly greater for the study-only condition ($M = 145.1$ sec) in comparison to time spent in the open-book test condition ($M = 100.1$ sec) and the closed-book test with feedback condition ($M = 95.9$ sec), ps $< .05$. The difference in time spent on the final comprehension tests for the open-book and closed-book test with feedback conditions was not significant, $p > .05$. There were no significant differences in time spent on final transfer tests for the study-only ($M = 108.4$ sec), open-book test ($M = 109.4$ sec), and closed-book test with feedback ($M = 113.1$ sec) conditions, $p > .05$.

Although subjects spent an additional minute during the closed-book test with feedback condition in the first session, final test performance for the closed-book test with feedback and open-book test conditions was equivalent. 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 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 subjects spent the least amount of time during the first session, the study-only condition, also produced the smallest benefit (in terms of delayed comprehension and transfer test performance) for the second session.

Discussion

Similar to the results from Agarwal et al., significant testing effects were found such that final comprehension and transfer test performance following initial open-book

and closed-book tests (with feedback) was greater than final performance for the study-only condition. Unfortunately, differences between the open-book and closed-book test with feedback conditions did not emerge on either the final comprehension tests or the final transfer tests. Surprisingly, performance on transfer questions (M = .67, collapsed over learning condition) 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. However, because subjects could rely on familiarity while completing the final multiple-choice tests (Jacoby, 1991; Kang et al., 2007), 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) may be the result of item differences.

Experiment 2

The results from Experiment 1 provide additional evidence that testing effects can be obtained with open-book and closed-book tests. These results also replicated findings from Agarwal et al. (2008) that initial open-book and closed-book tests with feedback do not produce different levels of performance on delayed retention tests. One point of departure between Agarwal et al. (2008) and Experiment 1 from typical educational settings is that students typically know the type of quiz or exam to expect with regards to open-book vs. closed-book format. In Agarwal et al. (2008) and Experiment 1 of this study, subjects were not made aware of the type of initial or final test to expect. Subjects

may have studied the passages in preparation for closed-book tests, since closed-book tests are more prevalent in both classroom and research settings (Feldhusen, 1961; Theophilides & Koutselini, 2000). If subjects used specific study strategies in preparation for closed-book tests, it is possible that they did not need to rely on the passage during the unexpected open-book test as much as they would have if they had expected the open-book test. Thus, we should not be surprised by the similar final test performance following initial open-book and closed-book tests in Agarwal et al. and Experiment 1 if subjects used similar encoding and retrieval strategies during both kinds of initial tests.

Prior findings of greater initial performance on open-book tests in comparison to closed-book tests may be caused by performance in the open-book test condition being measured *while* feedback was accessible, whereas performance in the closed-book test condition was measured *before* feedback was accessible. Considering that subjects accurately self-graded 77% of items following the closed-book test in Experiment 1, we can expect a high level of performance on a test immediately following the closed-book test with feedback condition, similar to initial performance in the open-book test condition in Experiment 1.

Furthermore, differences in initial performance do not imply that students used different study or retrieval strategies during the two tests; rather, differences in performance may have been an artifact of the timing of feedback. In Experiment 2, a questionnaire administered at the end of Session 2 provided an opportunity to examine subjects' study and retrieval strategies during the initial tests. In addition, the design of Experiment 2 was intended to mimic classroom settings by informing students of the type

of test to expect, as well as by allowing subjects to use any self-paced study strategy they prefer, during the initial tests.

In an attempt to investigate students' documented use of different study and retrieval strategies for open-book and closed-book tests in typical classroom settings (Feldhusen, 1961; Theophilides & Koutselini, 2000), we examined the effect of test expectancy instructions on initial test performance, long-term retention, and transfer of passage material in Experiment 2. Does test expectancy induce different encoding/studying strategies for open-book vs. closed-book tests? Does the type of encoding/studying before the initial test influence the type of retrieval strategies used, and subsequently affect delayed retention and transfer of information?

Experiment 2 included two design components intended to address methodological concerns articulated in the test expectancy literature (Neely & Balota, 1982; Schmidt, 1980). First, subjects received four initial passages and practice tests, two closed-book tests and two open-book tests, in order to equate encoding and retrieval practice with both tests, as well as to equate buildup of proactive interference. Second, the three critical test expectancy instructions (open-book, closed-book, and nonspecific) were manipulated between subjects and a substantial cash award was offered in order to avoid motivational differences at retrieval across the three groups (Zaromb, 2003).

Before reading each of the four practice passages, subjects were told to expect either an immediate open-book test or an immediate closed-book test (order was counterbalanced), and the initial test expected was always consistent with the initial test received immediately after studying. Subjects were not informed that retention for the practice passages would be tested in two days. Before the two critical passages, subjects

were told to expect a final open-book test, a final closed-book test, or a test in general (nonspecific expectancy instructions) during the second session. Thus for the critical passages, subjects did not expect (nor receive) immediate tests. Final tests, however, were always administered in a closed-book format during the second session in order to measure delayed retention. Closed-book tests in Experiment 2 were *not* followed by feedback in order to evaluate the effects of test expectancy, not feedback, on final retention.

The main comparison of interest in Experiment 2 was how open-book vs. closed-book test expectancy instructions before studying the two critical passages would influence final retention after two days. 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 hypothesized that the type of final test expected would influence final 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 final comprehension, transfer, and factual test performance.

A secondary interest of Experiment 2 was how students would study following the nonspecific expectancy instructions, and subsequently, how students' self-selected encoding/study strategies would influence final retention after two days. Upon completion of the entire experiment, subjects in the nonspecific expectancy group were asked to report which kind of final test they expected (i.e., studied for), open-book or closed-book. In general, we hypothesized that students reporting study strategies for an open-book test would have similar final retention performance to that of the open-book

expectancy group, and students reporting study strategies for a closed-book test would have similar final performance to that of the closed-book expectancy group. Based on previous literature, however, we predicted that a majority of students in the nonspecific expectancy group would expect a final closed-book test, and thus have similar final test performance to that of the closed-book test expectancy group.

Finally, we hypothesized that test expectancy instructions for the practice passages might influence initial passage reading and test performance, but initial performance on the two practice open-book tests was still expected to be greater than initial performance on the two practice closed-book tests due to accessibility to the passage, consistent with previous findings. Because initial test performance was expected to be similar in Experiment 2 to that in Experiment 1, and because subjects did not expect delayed tests on these passages, we predicted similar final test results for the practice open-book and closed-book test conditions, consistent with previous findings (Agarwal, et al., 2008).

Method

Participants. One hundred eight subjects were recruited from the Washington University in St. Louis Department of Psychology human subject pool. Subjects 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. During Session 1, subjects participated in a practice phase followed by a critical phase. During the practice phase, subjects read four passages and completed four multiple-choice GRE comprehension tests (alternating between reading and testing). For two of the practice passages, subjects were told to expect an immediate closed-book test,

and for two of the practice passages, subjects were told to expect an immediate open-book test. After reading a passage, subjects received an immediate test under conditions consistent with test expectancy (closed-book or open-book).

Next, during the critical phase, subjects received closed-book, open-book, or nonspecific instructions for the final test session. Then, all subjects read two critical passages. During Session 2 after a two-day delay, subjects were asked to complete final (closed-book) short answer transfer questions, short answer factual questions, and repeated multiple-choice comprehension questions on all six passages, followed by a questionnaire.

The 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 expectancy instructions during the practice phase followed two orders: 1) closed, open, closed, open, or 2) open, closed, open, closed. Thirty-six subjects were randomly assigned to each of the three between-subject critical test expectancy conditions, three subjects in each of twelve (6 passage by 2 practice test expectancy) orders.

Materials. The six passages and multiple-choice comprehension tests used in Experiment 2 were identical to those used in Experiment 1. Session 2 transfer test questions were identical to those used in Experiment 1, except multiple-choice alternatives were not presented; i.e., transfer questions were short answer. The experimenter constructed six factual short answer questions for each passage (see Appendix B for example materials from Experiment 2).

Procedure. Subjects were tested individually or in small groups. In Session 1, subjects were instructed that they would read several prose passages and take multiple-

choice tests. Subjects were also instructed that the top scoring participants would receive a $20 cash reward. While all passages and tests were presented in a paper-and-pencil format, subjects were seated at a computer and use an E-Prime 1.0 program (Schneider, Eschman, & Zuccoloto, 2002), which provided instructions and recorded time spent on each phase of the experiment.

During Session 1, which lasted approximately 60 minutes, subjects 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, subjects 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." For the open-book practice test condition, subjects 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." For the four initial practice tests, expectancy instructions and actual test received always matched (e.g., when a subject expected an open-book test, the passage was followed by an open-book test). Subjects were not, however, informed that these were practice tests, nor were they informed that the fifth and sixth passages were "critical," in order to maintain equal levels of high motivation across all passages and tests. For the two closed-book practice tests, subjects read the passage and then completed a multiple-choice comprehension test without viewing the passage; feedback was not provided. For the two open-book practice tests, subjects read the passage and

then were allowed to view the passage while completing a multiple-choice comprehension test.

Following the practice phase, subjects received closed-book, open-book, or nonspecific expectancy instructions for the final test session, and then read two passages during the critical phase of Session 1. For the closed-book test expectancy condition, subjects 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." For the open-book test expectancy condition, subjects 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." For the nonspecific expectancy condition, subjects 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."

All study and test periods were self-paced, although a maximum of four minutes (based on response times recorded in Experiment 1) per period was imposed in order to mimic time constraints in typical classroom settings. Before each study and test period, subjects were reminded of the four-minute time limit. During a study period, the computer instructed subjects to take a passage from a blue folder, read it at their own

pace, place it face down in a red folder when they were finished, and push spacebar on the keyboard to move on to the test (the E-Prime program recorded time spent studying). During a test, subjects were asked to take the corresponding test from the same blue folder (passages and tests were pre-arranged according to counterbalance order for each subject), circle a multiple-choice alternative for every question or write in their answer, place the test face down in the red folder when they completed the test, and push spacebar to move on to the next passage (recording time during testing). The experimenter observed compliance with all instructions provided.

Session 2 occurred two days after Session 1 and lasted approximately 60 minutes. Subjects completed final (closed-book) short answer transfer questions, short answer factual questions, and repeated multiple-choice comprehension questions for each passage; feedback was not provided. Tests were blocked by passage, such that when a subject finished the transfer, factual, and comprehension tests for one passage, they moved on to the transfer, factual, and comprehension tests for the next passage. Tests for critical passages were always followed by tests for the remaining four practice passages, always in the order in which subjects first encountered the passages during Session 1.

Finally, subjects were asked to complete a short questionnaire about study/test strategies and prior experience with closed-book and open-book tests (see Appendix C; question 4 adapted from Farr, Pritchard, & Smitten, 1990). Subjects in the nonspecific expectancy group were asked, at the end of the experiment, which kind of test they actually expected during Session 2. All subjects were debriefed and thanked for their time.

Results

Initial Test Performance. Initial closed-book and open-book practice test performance is shown in Table 5. Replicating Experiment 1, initial test performance was significantly greater on open-book tests (M = .68) in comparison to closed-book tests (M = .62), $F(1, 105) = 7.76$, $\eta_p^2 = .07$. As expected, there was no interaction between initial test performance and the three between-subject groups (open-book expectancy, closed-book expectancy, and nonspecific expectancy), $p > .05$, as the critical manipulation did not occur until after the initial practice tests.

Final Test Performance. Final test performance for the closed-book and open-book practice test conditions is shown in Table 5. Final 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). In fact, comprehension performance for the closed-book practice test condition slightly increased from 62% to 63% between the first and second sessions, even though feedback was not provided.

Similar patterns held for the final 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) and the final short answer fact test (M = .20 for passages initially tested an open-book format, and M = .22 for passages tested in a closed-book format). A 2 (practice test condition: open-book, closed-book) x 3 (final test type: comprehension, transfer, fact) ANOVA revealed only a significant effect of test type on final test performance, $F(2, 214) = 487.70$, $\eta_p^2 = .82$. These results confirm our hypothesis that type of initial test received, open-book or closed-book, does not influence final retention.

Table 5

Initial and final test performance (proportion correct) for practice test conditions in Experiment 2

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

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

Regarding final test performance for the three between-subject test expectancy groups (open-book expectancy, closed-book expectancy, and nonspecific expectancy), the nonspecific expectancy group (M = .40) showed similar performance to the closed-book test expectancy group (M = .39) across the three final tests (multiple-choice comprehension, short answer transfer, and short answer fact), and these two groups showed greater final test performance than the open-book expectancy group (M = .33).

Because the main comparison of interest in Experiment 2 was how open-book vs. closed-book test expectancy instructions would influence final retention after two days, data for the nonspecific expectancy group was subdivided into either the open-book or the closed-book expectancy group, depending on subjects' self-reported expectancy and study strategies used. Of the 36 subjects in the nonspecific expectancy group, 25 reported closed-book test expectancy, three reported open-book test expectancy, and eight did not respond. Thus, the recoded closed-book expectancy group included 61 subjects, the open-book expectancy group included 39 subjects, and the subjects who did not respond were dropped from further analyses. This distribution of responses (i.e., 69% of subjects self-reported expecting a closed-book final test in the nonspecific expectancy group) and similar levels of performance between the nonspecific and closed-book expectancy groups confirm our hypothesis that, in the absence of specific test instructions, a majority of students expect and subsequently study for closed-book tests.

Final test performance for the two critical conditions (open-book expectancy and closed-book expectancy) is shown in Table 6. A 2 (expectancy: closed-book, open-book) x 3 (final test type: comprehension, transfer, fact) mixed factors ANOVA confirmed a significant effect of expectancy, $F(1, 98) = 5.66$, $\eta_p^2 = .06$, and a significant effect of

Table 6

Final test performance (proportion correct) for critical expectancy conditions in Experiment 2

	Delayed Comprehension Test	Delayed Transfer Test	Delayed Fact Test	Delayed Average (by condition)
Expected closed-book final test (N = 61)	.59 (.02)	.38 (.03)	.23 (.02)	.40 (.02)
Expected open-book final test (N = 39)	.53 (.03)	.29 (.03)	.18 (.03)	.33 (.02)
Average (by test)	.56 (.02)	.33 (.02)	.20 (.02)	

Note. Data for the nonspecific expectancy group was subdivided into either the open-book or the closed-book expectancy group, depending on subjects' self-reported expectancy and study strategies used. Delayed tests occurred two days after initial comprehension tests. Standard errors of the mean are displayed in parentheses.

final test type, $F(2, 196) = 170.91$, $\eta_p^2 = .64$. As can be seen in Table 6, overall final test performance was greater for the closed-book expectancy group ($M = .40$) than the open-book expectancy group ($M = .33$). In addition, performance was greatest on the final multiple-choice comprehension test ($M = .56$), followed by performance on the short answer transfer ($M = .33$) and fact tests ($M = .20$), respectively.

Given our initial hypothesis that final performance should be greater following closed-book test expectancy than open-book test expectancy, we conduced planned one-tailed t-tests. These one-tailed t-tests confirmed that final comprehension test performance for the closed-book expectancy group ($M = .59$) was greater than final comprehension performance for the open-book expectancy group ($M = .53$), $t(98) = 1.68$, $d = .34$; final transfer test performance was greater for the closed-book expectancy group (.38 vs. .29) , $t(98) = 2.35$, $d = .49$; and this difference was marginally significant for final fact test performance (.23 vs. .18), $t(98) = 1.37$, $d = .28$, $p = .09$.

Response Times. Response times for Session 1 are shown in Table 7. Due to computer error during Session 1, response times for some passages and tests were not recorded, thus the number of subjects included in each analysis vary. During the first session, there was no significant difference in time spent reading passages for the open-book ($M = 128.7$ sec) and closed-book ($M = 136.8$ sec) practice test conditions, $p > .05$ (N = 103), even though subjects were informed of the type of test to expect immediately after reading the passage. Time spent completing the open-book ($M = 167.3$ sec) practice tests was significantly greater than time spent completing the closed-book ($M = 149.6$ sec) practice tests, $F(1, 79) = 5.49$, $\eta_p^2 = .07$ (N = 80), suggesting that subjects made use of the available passage while completing the open-book practice tests.

Table 7

Response times (seconds) in Experiment 2

	Session 1		Session 2	
	Reading	Initial Comprehension Test	Delayed Comprehension Test	Delayed Transfer and Fact Test
Practice closed-book tests	136.4 (104)	148.2 (107)	109.6	265.7
Practice open-book tests	129.7 (106)	168.1 (81)	106.9	270.2
Expected closed-book final test (N = 61)	167.6 (52)		160.7	401.0
Expected open-book final test (N = 39)	128.6 (38)		173.9	373.2

Note. Data for the nonspecific expectancy group was subdivided into either the open-book or the closed-book expectancy group, depending on subjects' self-reported expectancy and study strategies used. Session 2 occurred two days after Session 1. Due to computer error during Session 1, response times for some passages and tests were not recorded, thus the number of subjects included in each average are displayed in parentheses.

Regarding the two critical between-subject expectancy groups (again, subjects in the nonspecific expectancy group were included in either the closed-book or open-book expectancy group, based on self-reported expectancy), a main effect of test expectancy on reading times during Session 1 was demonstrated, $F(1, 89) = 15.43$, $\eta_p^2 = .15$ (N = 90), indicating that subjects tailored their effort based on the final test expectancy instructions provided, spending significantly less time reading when expecting a final open-book test ($M = 128.6$ sec) than when expecting a final closed-book test ($M = 167.6$ sec).

Response times for Session 2 are also shown in Table 7. The E-Prime program used in this experiment collected total time spent on the short answer test, thus time spent on the individual transfer and fact tests are collapsed. A 2 (practice test condition: open-book, closed-book) x 2 (final test type: multiple-choice comprehension, short answer transfer/fact) ANOVA revealed only a significant effect of test type on response time, $F(1, 107) = 304.60$, $\eta_p^2 = .74$, such that time spent completing the short answer test ($M = 268.0$) was greater than time spent completing the multiple-choice test ($M = 108.3$).

A 2 (expectancy: closed-book, open-book) x 2 (final test type: multiple-choice comprehension, short answer transfer/fact) mixed factors ANOVA also revealed a significant effect of test type on response time, $F(1, 98) = 157.44$, $\eta_p^2 = .62$, such that time spent on completing the short answer test ($M = 387.1$ sec) was greater than time spent completing the multiple-choice test ($M = 167.3$ sec). Thus, for Session 2, type of initial test condition or test expectancy instructions did not influence final test taking time; the only significant difference revealed was that subjects took longer to complete short answer tests than multiple-choice tests.

Questionnaire Results. Of the 108 subjects who participated in this experiment, fifty-one subjects (47%) preferred open-book class examinations; thirty-three subjects (31%) preferred closed-book exams; and twenty-four subjects (22%) had no preference. The two most common reasons subjects provided for preferring an open-book exam was that they felt they were easier ($N = 16$) and that they favored using reference material during an exam ($N = 16$). The most commonly cited reason for preferring a closed-book exam ($N = 13$) was that they felt that closed-book exam material is typically easier than open-book exam material.

The most common strategy used during open-book tests as self-reported by subjects ($N = 50$) was reading the entire passage, reading/answering each question, and then checking the passage for correct answers. Twenty-seven subjects self-reported partially reading the passage before answering questions, 21 subjects reported reading all questions before reading the entire passage, six subjects reported alternating between reading/answering questions and searching the passage for answers, and four subjects reported other strategies.

At the end of the experiment, subjects in the nonspecific expectancy group were asked to report the kind of test they actually expected during Session 2. Of these 36 subjects, 69% ($N = 25$) expected a closed-book final test, 8% ($N = 3$) expected an open-book final test, and eight subjects did not respond. Considering these results and similar final test results for the nonspecific and closed-book expectancy groups, we confirm our prediction that, in the absence of specific test format instructions, subjects expect and subsequently study for closed-book tests.

Discussion

As predicted, a significant main effect of test expectancy on final comprehension performance was demonstrated, and the open-book test expectancy instructions produced the lowest level of final test performance on comprehension, transfer, and factual questions. Consistent with Experiment 1 and Agarwal et al. (2008), initial open-book and closed-book tests did not produce different levels of final test performance. Subjects adjusted their study time during Session 1 in accordance with final test expectations, which resulted in a similar pattern of final test performance during Session 2: subjects spent the most time reading passages when provided closed-book test expectancy instructions and subjects in this condition had the greatest final comprehension, transfer, and fact performance. It is important to note that overall delayed test performance following initial tests (M = .42) was greater than delayed test performance following test expectancy instructions (M = .37), 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 potential benefits of completing open-book and closed-book tests for enhancing subsequent learning. Experiment 1 replicated findings from Agarwal et al. (2008): significant testing effects were obtained following open-book and closed-book tests with feedback (in comparison to the study-only condition), though the two types of initial tests did not produce differential benefits for long-term retention or transfer of knowledge. In Experiment 2, the critical role of test expectancy regarding open-book and closed-book tests was examined. Initial retrieval practice during these two types of tests did not yield

differences in long-term retention or transfer (consistent with Experiment 1), but final test expectancy significantly influenced initial reading time and delayed test performance. Confirming our hypothesis, closed-book test expectancy resulted in greater final comprehension, transfer, and fact test performance than open-book test expectancy.

A topic of interest throughout this project was whether open-book vs. closed-book tests would differentially benefit subjects' ability to transfer knowledge from initial GRE comprehension questions to final transfer questions that required causal reasoning. In both Experiment 1 and Experiment 2, the type of initial test condition (open-book or closed-book) did not influence final transfer test performance during the second session. In Experiment 2, however, closed-book test expectancy increased final transfer test performance by 9% relative to open-book test expectancy. While some educators argue that open-book tests promote transfer (Jacobs & Chase, 1992), our findings are consistent with the constructivist theory that if a reader is not required to construct a meaningful situation model, as may be the case in open-book testing when a student can rely on the passage for information, the reader will not generate inferences (Graesser, et al., 1994). Still, future research should aim to examine whether students' retention or transfer for other kinds of materials can benefit from open-book testing more than in the current experiments.

A second topic of interest in the current research was the potential influence of test expectancy on students' study behaviors and final test performance. In real-world educational settings, students are informed of the type of questions and exams (e.g., multiple-choice, short answer, essay, closed-book, open-book, take home, etc.) to expect during a course, but previous work on open-book and closed-book tests (e.g., Agarwal et

al., 2008) did not investigate this critical component of testing in applied settings. In Experiment 2, it was hypothesized that 1) in the absence of test expectancy, students study for closed-book tests, and 2) in the presence of test expectancy, students expend more effort while studying for a closed-book test than for an open-book test.

Confirming the first hypothesis, a majority of subjects in the nonspecific expectancy group later reported actually expecting a closed-book test and final test performance was similar for the nonspecific and closed-book expectancy groups. Confirming the second hypothesis, subjects in the open-book expectancy group spent the least amount of time reading during the first session and had the lowest level of performance across the delayed comprehension, transfer, and fact test in the second session. While these results require replication, preliminary conclusions can be drawn such that nonspecific and closed-book test expectancy instructions increase students' study time and subsequently enhance final test performance.

Although open-book and closed-book tests improved final retention in Experiment 1 more than simply reading a passage, the concept of desirable difficulty (Bjork, 1994) and the transfer appropriate processing framework (Bransford et al., 1979) do not apply to our consistent finding that long-term retention is equivalent following both types of tests. It was originally hypothesized that initial closed-book tests would produce greater final test performance than an initial open-book test based on both of these theories. Although encoding processes were "matched" when subjects received an initial closed-book test and a final closed-book test, this condition did not produce greater final performance than the "mismatch" condition (initial open-book test and a final closed-book test), inconsistent with the transfer appropriate processing framework. In

addition, while initial closed-book tests may be more difficult than initial open-book tests (i.e., based on increased reading time and a lower level of initial performance), initial closed-book tests did not enhance final performance more than open-book tests, contrary to the concept of desirable difficulty.

Instead of type of initial retrieval practice on a closed-book test, the *expectation* of a final closed-book test seems to be more potent for long-term learning. Perhaps students' prior experience with both open-book and closed-book tests (e.g., the SAT, a required college entrance exam, includes questions in open-book and closed-book formats; College Board, 2009) leads students to judge closed-book tests as more difficult, subsequently adapting study habits accordingly in preparation for the final test. Simply put, students' study habits may be based, in large part, on the perceived difficulty of a final test. It would be worthwhile to examine the combination of type of initial test followed by type of final test expected that is optimal for long-term retention. For the present, however, open-book vs. closed-book test expectancy instructions appear to drive differences in final retention more than open-book vs. closed-book initial retrieval practice, so it may be most beneficial for teachers to provide nonspecific or closed-book final test expectancy instructions. It is also recommended that teachers administer initial tests or pop quizzes, as initial tests improved long-term retention in both experiments, although the open-book vs. closed-book initial test format distinction may have little influence on final test performance.

In sum, these experiments demonstrate that test expectancy can influence study and test behaviors, as well as final retention, more than type of initial retrieval practice received. Students' expectations, test preparation methods, and strategy adoptions may

need to be considered carefully, particularly as interest in educationally relevant testing effect research continues to grow.

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.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bransford, J. D., Franks, J. J., Morris, C. D. & Stein, B. S. (1979). Some general constraints on learning and memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 331-354). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

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.

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

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

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

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.

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

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.

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. (in press). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New frontiers in applied memory*. Brighton, UK: Psychology Press.

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

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

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

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

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.

Tussing, L. (1951). A considération of the open book examination. *Educational and Psychological Measurement*, *11*, 597-602.

Zaromb, F. M. (2003). *Test expectancy and testing effects on memory for prose*. Unpublished manuscript, Washington University, St. Louis, MO.

Appendix A

Experiment 1 Materials

Passage

William Penn

Pennsylvania was the most successful of the proprietary colonies. Admiral Sir William Penn was a wealthy and respected friend of Charles II. His son, William, was an associate of George Fox, founder of the Society of Friends—a despised Quaker. When the senior Penn died, in 1670, his Quaker son inherited not only the friendship of the Crown but also an outstanding unpaid debt of some magnitude owed to his father by the King. As settlement, in 1681 he received a grant of land in America, called "Pennsylvania," which he decided to use as a refuge for his persecuted coreligionists. It was a princely domain, extending along the Delaware River from the 40th to the 43rd parallel. As Proprietor, Penn was both ruler and landlord. The restrictions on the grant were essentially the same as those imposed on the second Lord Baltimore: colonial laws had to be in harmony with those of England and had to be assented to by a representative assembly.

Penn lost little time advertising his grant and the terms on which he offered settlement. He promised religious freedom and virtually total self-government. More than 1,000 colonists arrived the first year, most of whom were Mennonites and Quakers. Penn himself arrived in 1682 at New Castle and spent the winter at Upland, a Swedish settlement on the Delaware that the English had taken over; he renamed it Chester. He founded a capital city a few miles upstream and named it Philadelphia—the City of Brotherly Love. Well situated and well planned, it grew rapidly. Within two years, it had

more than 600 houses, many of them handsome brick residences surrounded by lawns and gardens.

Shiploads of Quakers poured into the colony. By the summer of 1683, more than 3,000 settlers had arrived. Welsh, Germans, Scotch-Irish, Mennonites, Quakers, Jews, and Baptists mingled in a New World utopia. Not even the great Puritan migration had populated a colony so fast. Pennsylvania soon rivaled Massachusetts, New York, and Virginia. In part its prosperity was attributable to its splendid location and fertile soils, but even more to the proprietor's felicitous administration. In a series of laws—the Great Law and the First and Second Frames of Government—Penn created one of the most humane and progressive governments then in existence. It was characterized by broad principles of religious toleration, a well-organized bicameral legislature, and forward-looking penal code.

Another reason for the colony's growth was that, unlike the other colonies, it was not troubled by the Indians. Penn had bought their lands and made a series of peace treaties that were scrupulously fair and rigidly adhered to. For more than half a century, Indians and whites lived in Pennsylvania in peace. Quaker farmers, who were never armed, could leave their children with neighboring "savages" when they went into town for a visit.

By any measure, Penn's "Holy Experiment" was a magnificent success. Penn proved that a state could function smoothly on Quaker principles, without oaths, arms, or priests, and that these principles encouraged individual morality and freedom of conscience. Furthermore, ever a good businessman, he made a personal fortune while treating his subjects with unbending fairness and honesty.

Comprehension Questions

1. 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

2. The author mentions the "Holy Experiment" as an example of

    (A) English-Colonial collaboration

    (B) an early bicameral

    (C) a treaty with Indians

    (D) religious toleration

    (E) a reason for establishing a proprietary colony

3. How did Pennsylvania come into existence?

    (A) Penn found unexplored land and named it Pennsylvania

    (B) Penn received a grant of land in America from King Charles II

    (C) Penn’s father started Pennsylvania and gave it to him a few years later

    (D) Penn was granted a Swedish settlement and named the land Pennsylvania

    (E) Penn conquered an Indian settlement and named the land Pennsylvania

4. It can be inferred from the selection that

(A) all other colonies would have grown more rapidly if they had been organized in a manner similar to Pennsylvania

(B) all colonies should have been in harmony with the laws of England and had a representative assembly

(C) those colonies that were awards for service from the crown were better-administered

(D) the Pennsylvania Colony was the first colony to experience a tolerance for a number of nationalities and varied religious groups

(E) life with the Indians would have been much easier in other colonies if land had been purchased and treaties adhered to

5. The "Great Law" and the "First and Second Frames of Government"

(A) established Penn's political reputation

(B) created treaties with the Indians

(C) became the basis of a progressive republic form of government

(D) placed restrictions on immigration

(E) had to be overturned when they became inefficient

6. After the summer of 1683 the Pennsylvania colony could be referred to as

(A) a "melting-pot" colony

(B) a Quaker colony

(C) the largest American colony

(D) a Colonial Republic

(E) the first "democratic" colony

Transfer Questions

1. 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

2. Why was the "Holy Experiment" a magnificent success?

    (A) because the colony prospered without church sponsorship

    (B) because many different religions established themselves

    (C) because religions did not dictate the established laws

    (D) because Pennsylvania functioned smoothly without oaths, arms, or priests

    (E) because the Indians assimilated to the religious views of the Quakers

3. Why did William Penn receive a grant of land in America?

    (A) because of the death of an unknown wealthy relative

    (B) because he was a respected friend of the King

    (C) because of an unpaid debt owed to his father by King Charles II

    (D) because he was an associate of George Fox

    (E) because he wanted to begin a new colony for the Quakers

4. Why did Philadelphia grow rapidly?

    (A) because it was well situated and well planned

    (B) because it contained many natural resources

    (C) because many people settled in small spaces

(D) because many colonists came when they heard of the opportunities

(E) because it was considered safe

5. Why was Pennsylvania one of the most humane and progressive governments in existence at the time?

(A) because the people were not permitted to carry guns around

(B) because of equal rights for all citizens

(C) because Quakers believed in religious toleration

(D) because there was no death penalty instituted for crimes

(E) because of Penn's series of laws

6. Why might Pennsylvania be considered a "melting-pot" community?

(A) because of its bicameral legislature

(B) because the settlers came from different political parties of England

(C) because the colony did not have a caste system

(D) because the Indians did not trouble the settlers but acted rather peacefully

(E) because of the guarantee of freedoms and several religious sects in the colony

Appendix B

Experiment 2 Materials

Final Short Answer Transfer Test

William Penn

1. Why was Penn an opportunistic businessman?

_____

_____

2. Why was the "Holy Experiment" a magnificent success?

_____

_____

3. Why did William Penn receive a grant of land in America?

_____

_____

4. Why did Philadelphia grow rapidly?

_____

_____

5. Why was Pennsylvania one of the most humane and progressive governments in existence at the time?

_____

_____

6. Why might Pennsylvania be considered a "melting-pot" community?

_____

_____

Final Short Answer Fact Test

1. What was the full name of the king that provided Penn the grant of land?

2. In what year did Penn receive a grant of land in America?

3. How many settlers arrived in the first year of Pennsylvania?

4. What did Penn re-name a Swedish settlement on the Delaware River?

5. Aside from Quakers, what group of people formed the majority population during the first year in Pennsylvania?

6. According to the passage, how long (in years) did Indians and whites live peacefully in Pennsylvania?

Appendix C

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): _____

_____