Title
Multiple-choice testing can improve the retention of nontested related information

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Little, Jeri
Bjork, Elizabeth Ligon

Publication Date
2010

Peer reviewed

Multiple-choice testing can improve the retention of nontested related information

Jeri L. Little (jerilittle@ucla.edu)
Department of Psychology, 1285 Franz Hall, Box 951563
Los Angeles, CA 90024 USA

Elizabeth Ligon Bjork (elbjork@psych.ucla.edu)
Department of Psychology, 1285 Franz Hall, Box 951563
Los Angeles, CA 90024 USA

Abstract

Taking an initial test leads to improved performance on later tests for those previously tested questions. Whether prior testing improves one's ability to answer related questions, however, is less clear, with some results showing impairment for related information, an effect called retrieval-induced forgetting (RIF; e.g., Anderson, Bjork, & Bjork, 1994). Two experiments investigated the use of initial multiple-choice tests on the retention of previously studied prose passages, specifically on the retention of related, but initially nontested information. In both experiments, an incorrect alternative on the initial test served as the correct answer to a related question on the final test. Results demonstrated that the retention of related information can, indeed, be facilitated by initial multiple-choice tests (Experiment 1) and that this benefit is dependant upon using competitive incorrect alternatives (Experiment 2). We discuss how and why our results differ from previous work (e.g., RIF) and address possible educational applications.

Keywords: memory; testing effects; prose passages; RIF

Introduction

Testing is ubiquitous in education. In most cases, teachers use tests to assess how much a student has learned. Similarly, when students self-test (e.g., with flashcards or practice tests), they typically do so in order to assess their current mastery of the to-be-learned materials. Testing, however, can have other benefits that extend beyond evaluation because retrieval modifies memory so as to improve future recall (see Bjork, 1975).

Multiple-Choice Tests in Educational Contexts

Nowhere is the implementation of testing more widespread than in educational contexts, and in such contexts, the use of multiple-choice (MC) tests is very popular. Some concerns exist regarding their use, however. One concern is that MC tests might provide less opportunity for learning than do cued-recall (e.g., short answer) or free-recall (e.g., essay) tests. Indeed, some studies have shown that although initial MC, cued-recall, and free-recall tests all lead to better retention of the tested information, as compared to nontested information, retention of tested information is better after cued-recall or free-recall tests (e.g., Gay, 1980; Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Morrisette, & Derbish, 2007).

Perhaps the increased difficulty of answering a recall question correctly (versus a similar MC question) accounts for this difference; that is, retrieval, but not necessarily the recognition and selection of a correct answer, modifies memory (e.g., Bjork, 1975; McDaniel & Masson, 1985). We argue, however, that answering an MC question need not be just a matter of recognizing the correct alternative. In a well-constructed MC test, the test-taker likely recognizes most or all of the alternatives from previous study, but must decide whether or not that alternative is an appropriate answer to the question at hand (Sax & Collet, 1968; Whitten & Leonard, 1980). Often processes of discrimination and memory search are utilized as one thinks not only of which alternative is correct and why, but also of which alternatives are incorrect and why. Certain MC tests could, therefore, invoke a type of processing comparable to that invoked by recall tests (Whitten & Leonard, 1980).

Related Information

Although previous testing is clearly beneficial for retention of identical information, it is less clear whether testing might also benefit the retention of related, but initially nontested information. For example, if one reads a chapter about several U.S. presidents and then answers questions about some of those presidents, will information about the other presidents be strengthened as well? This issue seems particularly germane to the educational context where instructors would rarely ask the same questions on both a quiz and a later exam. In addition, instructors often give practice tests to students, with the intention of providing them with an idea of what the later exam will be like, while not providing them with the actual questions.

On the basis of previous research examining the effects of initial testing on the later recall of related information, one might expect the retention of such information to be impaired. To illustrate, using a retrieval-practice paradigm, Anderson, Bjork, and Bjork (1994) found that—after an initial study phase—testing or giving retrieval practice to some items from a given category improved their later recall, but impaired the recall of other items in that category that were not themselves tested, as compared to the recall of items from another category, none of which were tested—a phenomenon now known as retrieval-induced forgetting. Thus, it seems possible that by giving initial tests or practice questions, instructors could be inadvertently impairing their students' performance for related nontested questions that may appear on later exams.

Such retrieval-induced forgetting has been demonstrated for educational materials, including facts (Chan, 2009; Macrae & MacLeod, 1999); prose materials (Carroll, Cambell-Ratcliffe, Murnane, & Perfect, 2007); and even one's native language when words from a second language were practiced (Levy, McVeigh, Marful, & Anderson, 2007). Retrieval-induced forgetting has been argued to occur as the consequence of inhibitory processes needed to resolve competition among alternative responses to the same or similar cues (Anderson et al., 1994). Interestingly, it is argued that the processes that lead to forgetting are largely unconscious, as competitive alternatives need not be explicitly brought to mind for them to be suppressed. To the extent that related concepts are brought to mind, however, and can then be used to access the correct answer to a given question, related information might be facilitated.

Indeed, in recent work, Chan, McDermott, and Roediger (2006) developed question pairs such that answering one question on an initial test would encourage the spontaneous recall of information related to the second question that was then to be asked on the later test. Using these question pairs, Chan et al. found facilitation for related, but initially nontested information, although this result likely depended upon specific aspects of the procedure and materials in addition to the facilitative nature of the pairs (i.e., a 24-hr delay between the initial and final tests and integrated encoding of the to-be-learned information). In subsequent research, Chan (2009) demonstrated that although facilitation for these initially nontested, related items occurred at a 24-hr delay when the information had been learned in a prose context, forgetting occurred at a shorter delay when the information had been learned as an unordered series of facts. Importantly, in all of these studies that used short delays to final test, no facilitation was found for related information, even though the time spent on the initial test led to a greater amount of time-on-task—that is, time that the participant spent thinking about information from the tested passage.

In the present research, we tested whether MC tests might afford a benefit for related information that is not as easily afforded by cued-recall tests. Multiple-choice tests differ from free- and cued-recall tests in that they provide students with a set of related (and often competitive) concepts through which they can consciously search in selecting the correct answer, whereas cued-recall tests do not. For example, if given a cued-recall question about who served as the fourth president of the United States, although one may eventually recall the answer (i.e., *Madison*), in the process of doing so, other alternatives (e.g., *Adams, Jefferson*) may also become activated by the cue and compete for access and thus need to be suppressed in order to access *Madison,* according to inhibitory accounts of retrieval-induced forgetting. In contrast, if given an MC question with competitive alternatives provided (e.g., *Adams*, *Jefferson*), test-takers may be encouraged to consciously think about such competitors in selecting which president was the fourth (e.g., Adams and Jefferson held office prior to Madison, Jefferson was the third president, etc.), thereby strengthening information they spontaneously recall about these other presidents. Accordingly, MC tests (with competitive alternatives) might both reduce the possibility of retrieval-induced forgetting effects as well as encourage a type of spontaneous recall that later supports the enhanced recall of related, nontested information.

In Experiment 1, we explored this possibility by examining the effects of initial testing of some of the information presented in a prose passage on the later recall of related information using a variation of the retrieval-practice paradigm; specifically, we employed initial MC tests rather than cued-recall tests and then compared the recall of the previously tested items and related nontested items to that of control items from a passage not previously tested. We had two major questions in mind: (a) to what extent would the initial MC tests enhance the recall of previously tested information and (b) would the use of MC questions during initial testing enhance the recall of related information; that is, would the use of MC questions in the initial test allow related items to be facilitated instead of impaired—that is, escape retrieval-induced forgetting? In addition, we utilized a feedback manipulation to see whether being shown the correct answer after attempting to answer a question would affect later recall of both previously tested and related information. Although shown to improve recall of previously tested information, it is uncertain how feedback might affect recall of related information.

Experiment 1

Method

Participants A total of 112 students at the University of California, Los Angeles, participated for credit in an introductory psychology course.

Design We used a 2 (item type: previously tested, previously nontested related) x 2 (feedback: present, absent) within-subjects design plus an independent control group.

Materials Two passages were constructed, one about Saturn and one about Yellowstone National Park, and ten pairs of MC questions were created for each passage. The two questions in each pair were semantically related in that both questions tested the same topic (e.g., geysers) and had the same four alternatives (e.g., *Old Faithful*, *Steamboat Geyser*, *Castle Geyser,* and *Daisy Geyser*), but different correct answers (e.g., *What is the tallest geyser in Yellowstone National Park?* Answer: *Steamboat Geyser;* and, *What is the oldest geyser in Yellowstone National Park?* Answer: *Castle Geyser*). Questions were divided into two 10-item sets for a given passage, with the two questions from each pair randomly assigned to a different set.

Procedure All participants were given 10 min to read the first passage and were instructed to continue studying it if they finished early. Participants in the testing condition were then given an initial 10-item MC test (i.e., all items in one of the question sets for that passage) with questions presented one at a time on the computer. For a given test, all questions were either followed by feedback (feedback present) or not (feedback absent) after the participant provided an answer. Feedback entailed the entire question being re-presented, with the answer printed in red. Following study and test of the first passage, participants followed the same procedure for the second passage except that if feedback had been provided in the first MC test, then it was absent in the second test and vice versa.

Participants in the control condition received no tests; rather, they engaged in a non-verbal filler task (i.e., playing Tetris) following their study of each passage (for the same amount of time as would have been needed to take the test).

Finally, both tested and control participants received a final cued-recall test after a 5-min retention interval during which they played Tetris. Forty questions were presented one at a time on the computer screen; as cued-recall questions, they did not appear with any answer alternatives. For the tested condition, except for the absence of alternatives, half of the questions were identical to the MC questions (i.e., previously tested) and half were the nontested related items (i.e., the remaining questions from the two 10-item sets that had not appeared in the initial MC tests). Related questions were always tested before previously tested questions. For the control condition, all questions were previously nontested and served as a baseline. Topic (Passage) order, question set, and feedback (after Passage 1 or Passage 2) were counterbalanced.

Results and Discussion

Initial MC Test Performance Participants in the tested condition correctly answered an average of 70% (*SD* = 17%) of the questions on the initial MC tests.

Final Test Performance Final test performance is presented in Figure 1. As shown, we found evidence that taking an initial MC test improved the recall of both previously tested and previously nontested related information as compared to the control condition.[1]

Recall performance of participants in the tested condition was compared to the corresponding performance of participants in the nontested control condition via planned independent-samples *t* tests and, importantly, benefits were found for both types of questions. Specifically, these comparisons revealed that (a) previously tested questions given feedback (M = 65%, SE = 3%) and previously tested questions not given feedback (M = 51%, SE = 3%) were both answered correctly more often than the control questions (M = 29%, SE = 2%), $t(110) = 10.88$, $p < .001$, and $t(110) = 6.45$, $p < .001$, respectively; and (b) questions related to previously tested questions that had received feedback (M = 40%, SE = 3%) and questions related to previously tested questions that had not received feedback (M = 43%, SE = 3%) were both answered correctly more often than the control questions (M = 33%, SE = 2%), $t(110) = 2.10$, $p < .05$ and $t(110) = 3.10$, $p < .01$, respectively.

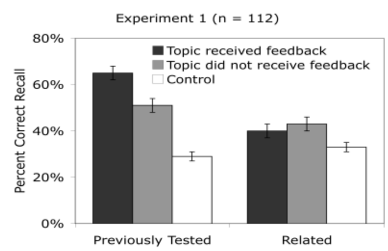

Figure 1: Correct recall percentages as a function of item and feedback type in Experiment 1. White bars show baseline recall for initially non-tested questions by control participants. Error bars represent +/- 1 SE.

To summarize, in Experiment 1, we found a generalized benefit of testing such that the answers to questions on a final cued-recall test were recalled more often when preceded by initial MC tests than when not. Most importantly, this benefit occurred even when the questions on the final cued-recall test were not identical, but only related to those on the initial MC tests, and even though answering such questions correctly on the final test involved recall of an answer that participants had needed to select against during the initial MC test. Thus, providing participants with practice on initial MC questions allowed related information not only to escape impairment but, indeed, to be enhanced. Although retrieval-induced forgetting is largely believed to occur as the result of the unconscious suppression of competitive alternatives, MC tests provide learners with the competitors and thus they can be consciously examined. For example, if students are given a set of alternatives that had all occurred in the required reading, as is the case for a MC question in our

[1] Overall, participants in the nontested control group correctly answered 31% (*SD* = 13%) of the questions on the final test, recalling marginally more answers in the first half of the test (M = 33%, SE = 2%) than in the second half (M = 29%, SE = 2%), $F(55) = 3.5$, $p = .07$, a finding consistent with previous accounts of output interference. Because of this marginal difference in performance for the first half and second half of the test, we compared recall for previously tested questions in the tested condition (which were always presented in the second half of the final test) with recall for questions in the control condition that were presented in the second half of the final test. Similarly, recall for nontested questions in the tested condition was compared with recall for the questions in the control condition that were presented in the first half of the test. This method of analysis provides a more conservative test of facilitation for related information.

experiment (e.g., *88 Earth days, 176 Earth days, 10 Earth hours*, and *30 Earth years* for the question: *How long does it take Saturn to revolve around the Sun?*), they could use these alternatives as a guide for searching their relevant knowledge set for the answer (e.g., 88 days and 176 days are wrong as they are related to Mercury; Saturn has a shorter day than Earth). Hence, even if students were unable to recall the answer to this particular question if asked in the format of a cued-recall question, if asked in the format of a MC question with possible alternatives provided, knowledge of related information presented in the passage might be utilized to reject incorrect alternatives; and, in this process, the student may spontaneously answer other related, but nontested questions. Indeed, we believe that such spontaneous retrievals may be the process by which the observed benefit for related but previously nontested items occurred in Experiment 1. For such a beneficial search process to be invoked, however, it would seem necessary that the incorrect choices be potential answers (i.e., competitive alternatives to the correct answer), thus requiring the student to select against them with the use of associated information from the passage. In contrast, without competitive alternatives, perhaps a benefit to related nontested information would not occur because the alternatives would not encourage this type of search strategy. We sought to explore this possibility in Experiment 2 by manipulating the competitiveness of the incorrect alternatives in the initial MC tests that followed the reading of prose passages.

Experiment 2

In Experiment 2, we tested whether the benefit of testing observed for related but previously nontested items in Experiment 1 arose from a type of search strategy engendered by the use of competitive alternatives in the initial MC tests, as described above. To do so, we manipulated the plausibility of the incorrect alternatives, hypothesizing that the more plausible the incorrect alternatives were as answers, the more competitive they would be and the more processing they would require in the attempt to reject them—processing that would likely involve retrieval of associated information from the passage and thus deeper processing of both the correct and the incorrect alternatives. Accordingly, we predicted that initial MC questions using more plausible incorrect alternatives would lead to a greater recall benefit for both previously tested information and previously nontested related information than would initial MC questions using less plausible incorrect alternatives.

Method

Participants A total of 28 students at the University of California, Los Angeles, participated for credit in an introductory psychology course.

Design We used a 2 (item type: previously tested, previously nontested related) X 2 (MC question type: competitive, non-competitive) within-subject design for the testing condition plus a control condition, with all participants serving in both conditions.

Materials Two passages were constructed, one about the Solar System and one about Ferrets, and eight question pairs were created for each passage. Related questions tested information from the same passage and the same type of information served as the correct answer for both questions (e.g., numbers, terms, proper names). To illustrate, the answers to two such questions (*How many inches long is an average ferret tail?* and *How many years ago were ferrets first domesticated, according to mitochondrial DNA evidence?*) were both numbers (i.e., *5* and *2500*, respectively).

To utilize a MC format for each of these questions, four incorrect alternatives were chosen from other information presented in the passage. Two incorrect alternatives were highly related to one question in the pair (and thus, plausible answers for it) and the other two alternatives were highly related to the other question (and thus, plausible answers for it). Thus, for a given pair, there were six alternatives (including the two correct answers). Because all of the alternatives for a given pair had the same type of answers (e.g., numbers), each of the six alternatives could be used in constructing two three-alternative MC questions for each question in these pairs. By manipulating which alternatives were used, we created a competitive and non-competitive version of each question. For example, in competitive versions, the incorrect alternatives were *7-10* and *20* for the first question and *1500* and *3500* for the second question. For the non-competitive versions, the incorrect alternatives were *1500* and *3500* for the first question and *7-10* and *20* for the second question.

Next, two new questions were constructed for each question-pair to serve as the nontested related questions on the final cued-recall test. As in Experiment 1, for these new questions, correct answers were previously incorrect alternatives on the MC questions. For example, although *7-10* was used as an incorrect alternative on the initial test, it was the correct answer to the question, *"How long do ferrets typically live?"* Similarly, *3500* was the correct answer to the question, "*According to archaeological evidence, how long ago were ferrets domesticated?"*

In summary, the six possible alternatives for each of the eight question-pairs were manipulated so as to make both of the three-alternative questions in each pair competitive or non-competitive for a given participant. On the initial MC test, all participants answered eight competitive questions and eight non-competitive questions. The final test included previously nontested questions for which previously incorrect alternatives (either competitive or non-competitive) were now the correct answers.

Procedure All participants were given 10 min to read the first of two passages and were instructed to continue studying the passage if they finished reading early. Next participants either took a test or engaged in a non-verbal filler task. When the passage was tested, participants were given an initial MC test with 16 questions (i.e., eight question pairs) for which they gave verbal responses. Questions appeared one at a time on a computer, and no feedback was given. When the passage served as the nontested control passage, participants played Tetris following its presentation for 3 min (the same time needed to take the test). If given a MC test after the first passage, then that participant engaged in the non-verbal filler task after the second passage and vice versa.

Finally, after a 4-min retention interval during which all participants played Tetris, a final 64-question cued-recall test was given. The 32 questions for the tested topic (previously tested and previously nontested related) and the 32 questions from the nontested control topic were presented on a computer screen, one-at-a-time, and participants gave a verbal response to each. Questions from the previously tested topic were always tested last. Topic (Passage) order, plausibility of alternatives, and testing (after Passage 1 or after Passage 2) were counterbalanced.

Results and Discussion

Initial MC Test Performance On the initial MC test, participants correctly answered more non-competitive questions (*M* = 86%, *SE* = 3%) than competitive questions (*M* = 66%, *SE* = 3%), $t(27) = 5.67$, $p < .001$, confirming that competitive alternatives make questions more difficult to answer correctly than do non-competitive alternatives.

Final Test Performance Final test performance is presented in Figure 2. For previously tested questions, correct answers to competitive questions (*M* = 37%, *SE* = 3%) were recalled marginally less often than were correct answers to non-competitive questions (*M* = 45%, *SE* = 4%), $t(27) = 1.76$, $p < .10$, a pattern consistent with the initial MC performance. For previously nontested questions from the same topic, however, the effect was in the opposite direction: correct answers that had previously been incorrect competitive alternatives (*M* = 47%, *SE* = 5%) were recalled more often than were correct answers that had previously been incorrect non-competitive alternatives (*M* = 36%, *SE* = 4%), $t(27) = 2.55$, $p < .05$, confirming our prediction that initial MC questions with competitive alternatives would lead to enhanced recall of related information as compared to initial MC questions with non-competitive alternatives.

When compared to control items (*M* = 27%, *SE* = 3%), answers to both previously tested competitive questions (*M* = 37%, *SE* = 3%) and previously tested non-competitive questions (*M* = 45%, *SE* = 4%) were facilitated, $t(27) = 3.10$, $p < .01$ and $t(27) = 4.54$, p < .001, respectively, demonstrating a testing effect. For previously nontested questions from the tested topic, those with answers that had previously been incorrect competitive alternatives (*M* = 47%, *SE* = 5%) were correctly answered more often than questions from the control passage (*M* = 36%, *SE* = 4%), $t(27) = 2.21$, $p < .05$, whereas those with answers that had been incorrect non-competitive alternatives (*M* = 36%, *SE* = 4%) were not, $t(27) = 0.1$, p > .05

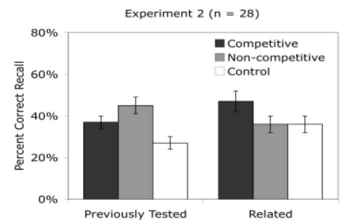

Figure 2: Correct recall percentages as a function of item type and competitiveness of MC alternatives on the initial MC test in Experiment 2. Error bars represent +/- 1 SE.

In Experiment 2, we manipulated the competitiveness of a given question by choosing incorrect alternatives that were either plausible or implausible answer choices to examine whether competitiveness of the alternatives was a critical factor in the facilitation of related information, and our results suggest this to be the case. Of concern, however, is whether the benefit we observed resulted from the increased processing of the incorrect alternatives as hypothesized, or simply occurred as an artifact of initial test performance. Because competitive questions were more difficult to answer than non-competitive ones, perhaps the benefit observed was merely a consequence of the participant being more likely—on the initial MC test—to select an incorrect competitive alternative than to select an incorrect non-competitive alternative, and then to recall that previously incorrect answer on the final test when given the related question (for which the answer might now be correct). For example, on the initial MC test, a participant might incorrectly choose *7-10* (instead of *5*) when given the question, "*How many inches long is an average ferret tail?*" If the participant then gives *7-10* as the correct answer for the question, "*How long do ferrets typically live?*" on the later test, one cannot be sure whether that participant is giving that answer believing it to be correct or giving that answer because it was chosen before and is now primed as an answer for all questions where it is plausible.

If such generalized strengthening of alternatives is the mechanism that leads to this effect, then participants should not demonstrate the pattern of results previously shown for related questions when recall is conditionalized upon answering the corresponding MC question correctly. A conditional analysis demonstrated, however, that marginally more answers to related questions were recalled correctly when those answers were previously incorrect competitive alternatives (*M* = 50%, *SE* = 4%) than when they were previously incorrect non-competitive alternatives (*M* = 41%, *SE* = 4%), $t(27) = 1.91$, $p = .07$. Thus, the possibility that this effect is driven by cases in which a participant chooses the incorrect answer and then carries it to a new question (where it then happens to be correct) seems unwarranted.

General Discussion

The present results imply that taking an initial MC test not only improves one's ability to recall *that* information on a later cued-recall test, but also improve one's ability to recall nontested, but related information on a later test—provided that the initial test utilizes incorrect alternatives that are competitive. Furthermore, although an MC question is often easier to answer than a comparable question in a cued-recall format (i.e., same question, without the choices), to the extent that the question has competitive alternatives, that question may invoke processes that are similar to those involved in recall (e.g., memory search, retrieval checks), thus leading to comparable benefits to the tested information. Moreover, use of MC questions may provide a way to insure that access to related nontested information is not impaired on a later test.

Educators may be concerned that the initial test provides participants with additional time to think about the tested topic, whereas no such additional time is allocated to the nontested control topic. Although a valid concern, our findings need to be viewed in the context of previous work using the retrieval-practice paradigm in which additional time is not allotted for nontested control materials and in which nontested information from a tested topic is rarely facilitated and, in fact, is typically impaired (e.g., Macrae & MacLeod, 1999; Carroll et al., 2007). Indeed, with a similar procedure and educational prose materials, but with an initial cued-recall test, Chan (2009) did not find facilitation for related information, even when the questions on the initial test were created to be facilitative for questions on the final test. One might thus argue that our finding of facilitation occurred because our MC questions—unlike cued-recall questions—exposed participants to the future answers for related questions (in the form of incorrect alternatives), thus providing shallow priming that leads to facilitation on the later test. Against such an argument, however, are the findings of Experiment 2 where alternatives were exposed in both competitive and non-competitive conditions and yet facilitation only occurred when alternatives were competitive, thus ruling out an explanation in terms of priming. Instead, our findings are consistent with the explanation that competitive MC questions lead to enhanced performance for nontested related information, owing to the deeper processing of the incorrect alternatives that they engender, as compared to processing engendered by noncompetitive MC questions.

We believe that the present results have implications for both instructors and students. Instructors can create quizzes and study guides that improve retention for both initially tested information as well as related information that is not itself tested. Students can benefit from tests by thinking about all of the alternatives—not only why a given answer is correct, but also why other answer choices are wrong.

Acknowledgments

We thank Ashley Kees for her valuable contributions in both the conception and implementation of this work and Robert Bjork for his valuable insights. Grant 29192G from the McDonnell Foundation supported this research.

References

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1063-1087.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* Hillsdale, NJ: Lawrence Erlbaum Associates.

Carroll, M., Campbell-Ratcliffe, J., Murnane, H., & Perfect, T. (2007). Retrieval-induced forgetting in educational contexts: Monitoring, expertise, text integration, and test format. *European Journal of Cognitive Psychology*, *19*, 580-606.

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

Gay, L. R. (1980). The comparable effects of MC versus short-answer tests on retention. *Journal of Educational Measurement*, *17*, 45-50.

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.

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second language acquisition. *Psychological Science, 18*, 19-34.

Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology*, *77*, 463-473.

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

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

Sax, G., & Collet, L. S. (1968). An empirical comparison of the effects of recall and MC tests on student achievement. *Journal of Educational Measurement, 5*, 169-173.

Whitten, W. B., & Leonard, J. M. (1980). Learning from tests: Facilitation of delayed recall by initial recognition alternatives. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 127-134.