The Persisting Benefits of Using Multiple-Choice Tests as Learning Events

Jeri L. Little (jerilittle@wustl.edu)
Department of Psychology, Washington University in St. Louis
Campus Box 1125, One Brookings Drive, St. Louis, MO 63130-4899

Elizabeth Ligon Bjork (elbjork@psych.ucla.edu)
Department of Psychology, University of California, Los Angeles
1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563

Abstract

Taking a test tends to improve the retention of the tested information. Additionally, taking a test often influences the later retention of non-tested information, provided such information is related to the tested information in a specific manner. To illustrate, recent research has demonstrated that multiple-choice tests containing competitive alternatives can improve retention of both tested and non-tested information pertaining to such incorrect alternatives at least over a short delay. The present research investigated whether such improvements in retention would persist with a delay more likely to occur in educational contexts (i.e., 48 hr). Taking an initial multiple-choice test improved retention more than a comparable cued-recall test—for both previously tested and related information—and over both short and long delays. Moreover, misinformation effects seen for the multiple-choice test at the short delay were reduced. These results thus have important implications for the use of multiple-choice tests as learning opportunities.

Keywords: Testing; test effects; multiple-choice; retrieval-induced forgetting, retrieval-induced facilitation

Introduction

Taking a test does more than assess one's knowledge: It can also improve one's long-term retention of the tested information. Not all tests, however, are equally beneficial in this manner. For example, more open-ended tests (e.g., cued-recall), in general, have been shown to improve long-term retention more than multiple-choice tests (see meta-analyses by Anderson & Biddle, 1975; Hamaker, 1986). Moreover, some research has shown that taking multiple-choice tests can actually harm later performance on open-ended tests by exposing the test takers to misinformation in the form of incorrect, but plausible, answer choices, with the consequence of such information sometimes being intruded as incorrect responses to later cued-recall questions—findings referred to as misinformation effects (e.g., Roediger & Marsh, 2005). Thus, multiple-choice tests are often accused of not only being less effective for learning than are more open-ended types of tests, but also to risk negative misinformation effects, bringing their use as learning tools in educational contexts into question.

Although, as indicated, multiple-choice tests can produce misinformation effects on later open-ended tests, research, in general, has shown that the positive effects of multiple-choice testing (as compared to no testing) outweigh any such negative effects they engender. Furthermore, recent research has indicated that multiple-choice testing may have a benefit for later overall performance that is not produced by the taking of prior cued-recall tests: namely, not only improving the retention of the information tested but also that of non-tested, but related, information—an outcome that would seem to be particularly desirable in educational contexts. When, for example, instructors give quizzes or practice tests to be followed later in the course by a more comprehensive exam, they are likely to ask questions about related information instead of (or in addition to) questions about the specific information tested earlier. Thus, it seems critical that the use of prior testing in educational contexts should help to improve the retention of both types of information.

With respect to the question of how retrieving some information affects the later retention of non-tested related information, previous research has demonstrated that the effects can be negative—that is, answering a cued-recall question can lead to impaired recall of competitive related information on a later test, a finding referred to as retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994). Even in cases in which retrieval-induced forgetting does not occur, however, the retention of non-tested related information is rarely facilitated as a consequence of a cued-recall test, especially when such non-tested information has a competitive relationship with the tested information. (For exceptions to this finding when tested and non-tested related questions were created to be facilitative, see Chan, McDermott, & Roediger, 2006, and Chan, 2009.)

In contrast to such findings for cued-recall tests, however, Little, Bjork, Bjork, and Angello (in press) recently demonstrated that—in addition to improving retention of previously tested information—taking an initial multiple-choice test, as compared to taking an initial cued-recall test, can improve retention of related information when the answers to the competitive related questions have occurred as incorrect alternatives in the initial multiple-choice test. Although these findings have clear implications for educational practice, particularly how an instructor might construct a practice exam, the procedure used by Little et al. employed a delay that is considered to lack educational realism (e.g., 5 min). Thus, a primary goal of the present research was to assess whether the demonstrated retention benefits (particularly for competitive related information) occurring as a consequence of taking multiple-choice tests—compared to cued-recall tests—would persist at more educationally realistic delays.

Retention of Non-tested Related Information

Although the effect of testing on the retention of both tested and related information was thoroughly explored in the 1960s and 70s (e.g., see literature on the use of adjunct questions, Anderson & Biddle, 1975; Frase, 1971; Hamaker, 1986; Rothkopf, 1970), interest in this topic has recently been renewed, not only because of the implications for educational practice, but also, in light of the finding of retrieval-induced forgetting (Anderson et al., 1994), to gain a better understanding of the circumstances under which an initial test might improve or impair retention of related information (e.g., Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan et al., 2006; Chan, 2009; Little, Storm, & Bjork, 2011).

The research by Little et al. (in press), in which they directly compared the effects on later recall of taking an initial cued-recall test versus a multiple-choice test following study of a passage, provides some insight into this issue. Because retrieval-induced forgetting seems to depend upon the occurrence of a competitive relationship between tested and non-tested related information (see, e.g., Storm, 2010), Little et al. specifically constructed related question pairs so that one question in the pair would have a competitive relationship with the other question in the pair. For example, given a passage about Saturn, one question would ask the participant "how long it takes Saturn to revolve around the Sun" to which the correct answer is "30 Earth years," and the other question would ask "how long it takes Saturn to rotate on its axis" to which the correct answer is "10 Earth hours." On the initial multiple-choice tests, the answer to the related question appeared as one of the incorrect alternatives (e.g., *How long does it take Saturn to rotate on its axis? a. 10 Earth hours, b. 88 Earth days, c. 176 Earth days, d. 30 Earth years*).

Little et al. hypothesized that the presence of competitive alternatives as answer choices on the multiple-choice question would induce students to think deeply about the question and alternatives—perhaps not only recalling information about why the correct alternative is right, but also why the incorrect alternatives are wrong—with this spontaneous recall of information pertaining to incorrect alternatives serving as a learning event for that information and potentially improving its future recall. In contrast, they hypothesized that a cued-recall test would not afford the opportunity for such beneficial processing of related competitive information; thereby, perhaps making it subject to retrieval-induced forgetting. And, indeed, their findings supported this type of analysis: Whereas the taking of an initial cued-recall test following study of the passage led to impaired later recall of non-tested competitive information (Exp. 1), the taking of a multiple-choice test after reading the passage not only spared such information from forgetting (Exp. 1), but facilitated its retention (Exp. 2).

Although one might wonder if the facilitated retention of related information observed by Little et al. (in press) might have occurred owing simply to the exposure of those answers as incorrect alternatives during the initial multiple-choice test, a study by Little and Bjork (2010) provides evidence that the benefit for the retention of related information would not have occurred for this reason. Little and Bjork manipulated whether the incorrect alternatives were competitive or non-competitive for a given question, with the expectation that alternatives would need to be competitive for facilitation to occur. To illustrate, students answered questions (e.g., *Which outer planet was discovered by mathematical predictions rather than direct observation?* Answer: *Neptune)* with competitive alternatives (e.g., *Neptune, Uranus, Saturn*) or with non-competitive alternatives (e.g., *Neptune, Mercury, Mars*), and they were better able to answer a later related question (for which *Uranus* or *Mercury,* respectively, were the correct answers) when the answer to that related question had served as a competitive alternative versus as a non-competitive alternative on the initial multiple-choice test. Thus, simple prior exposure to the incorrect alternatives cannot explain the observed results: Competitive alternatives seem to trigger the retrieval processes that support the improved retention of information pertaining to those incorrect alternatives.

The results of the experiment by Little and Bjork (2010), although theoretically informative, do not address the question of concern in the present research: namely, whether such effects persist over educationally realistic durations. On the one hand, the pattern of their results suggests that the presence of competitive alternatives on the initial multiple-choice test induces learners to engage in both deep processing of the choices and spontaneous recall of information pertaining to those choices, both of which might promote long-term retention. On the other hand, the pattern of results observed by Little et al. (in press, Exp. 1) might not persist. Their observation of a benefit of multiple-choice testing over cued-recall testing might reflect, in part, a temporary impairment in the recall of information related to the cued-recall questions, as other research has suggested that such retrieval-induced forgetting does not persist (Chan, 2009; MacLeod & Macrae, 2001). Chan, for example, found that although forgetting occurred as a consequence of an initial cued-recall test when the retention interval was short (i.e., 20 min), forgetting was not observed when the interval was long (i.e., 24 hrs). Furthermore, when forgetting did not occur at a short delay, facilitation emerged at a longer delay. Thus, it is possible that Little et al.'s observed pattern of a benefit of multiple-choice testing over cued-recall testing would be eliminated when retention is assessed at a delay longer than 5-20 minutes.

Retention of Previously Tested Information

In addition to assessing the effect of initial testing on the later recall of related information, we were also interested in its effect on previously tested information. Specifically, we also wanted to assess whether the benefit of multiple-choice testing compared to cued-recall testing observed by Little et al. (Exp. 1) would persist over a long delay. If, as has been suggested (e.g., Foos & Fischer, 1988), cued-recall tests involve deeper processing than do multiple-choice tests, then such an outcome would seem unlikely; that is, improved retention of the tested information would seem more likely to persist as a consequence of a cued-recall test than as a consequence of a multiple-choice test (see, e.g., Roediger & Karpicke, 2006), which might lead to a reversal in the direction of the effect.

Misinformation: A Negative Consequence of Multiple-choice Testing

Finally, multiple-choice tests can result in misinformation effects (Roediger & Marsh, 2005) and seem particularly prone to do so when feedback is not provided (Butler & Roediger, 2008). Thus, another goal of the present research was to assess the influence of a longer retention interval on the prevalence of misinformation effects.

The Present Experiment

In summary, Little et al. (in press) claimed that multiple-choice tests should be exonerated—at least from the criticisms involving their use in learning, but they used a retention interval that has limited application for educational purposes. In the present experiment, we assessed whether their observed benefits of taking a multiple-choice test over taking a cued-recall test or no test at all would persist over a longer retention interval. Additionally, we assessed the occurrence of misinformation at these two retention intervals.

Method

Participants and Design

Seventy-two undergraduates at the University of California, Los Angeles, who were all recruited for a two-session study, participated for either partial course credit or payment. Each participant read three passages, followed immediately by two initial tests: a cued-recall test for one passage and a multiple-choice test for another passage (the remaining passage was not tested initially and served as the control condition). After a delay (either 5-min or 48-hr), all participants took a final test that included previously tested questions and previously non-tested related questions (pertaining to multiple-choice questions for one passage and cued-recall questions for another passage), and non-tested questions from the non-tested passage.

Materials

The same materials as used by Little and Bjork (2011, Exp. 3) were employed: Three passages of approximately 800 words each about Saturn, Yellowstone National Park, and stimulant drugs with ten pairs of multiple-choice questions for each passage. The two multiple-choice questions in each pair were semantically related in that both 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*). For each passage, the questions in each pair were randomly assigned to a different 10-item set. Cued-recall questions were the same as the multiple-choice ones, but without alternatives provided.

Procedure

All participants were given 6 min to read each of the three passages in immediate succession, being told to spend the full time reading and studying. After this initial study phase, they were given a 10-item multiple-choice test for one passage and a 10-item cued-recall test for another passage (with each test containing all the items in one of the question sets for that passage). Questions were shown on the computer screen one at a time for 24 s, and participants typed in their responses, being instructed to spend the full time thinking about the question and their answer. Each question was answered twice (as is common in studies of retrieval-induced forgetting) such that participants answered each of the ten questions once, and then answered each question again, but in a different order.

After taking the initial multiple-choice and cued-recall tests, all participants engaged in a non-verbal 5-min distractor task of maze completion. Then, participants randomly assigned to the 5-min delay condition immediately took the final test; those randomly assigned to the 48-hr delay condition were excused, with the instruction to return 48 hrs later.

On the final test, participants answered 60 cued-recall questions, with the questions presented one at a time on the computer screen. Of the 20 final-test questions associated with the passage that received an initial cued-recall test, half were identical to the 10 cued-recall questions initially asked about that tested passage, while the other half consisted of the related questions that had not been presented on the initial cued-recall test (i.e., the remaining questions from the set of 10-paired questions constructed for that passage). Of the 20 final-test questions associated with the passage that received an initial multiple-choice test, half were identical to the 10 multiple-choice questions initially asked about the tested passage (except they were now asked as cued-recall questions without alternatives presented), while the other half consisted of the related questions that had not been presented on the initial multiple-choice test (i.e., the remaining questions from the set of 10-paired questions constructed for that passage). Of the 20 final-test questions related to the non-tested passage, all were previously non-tested and served as baseline control items.

As the later recall of non-tested related information was most crucial to our research questions, the non-tested questions from the tested passages were tested in the first half of the test, along with half of the non-tested control questions, to which their performance would be compared. Previously tested questions were tested in the second half of the test, along with the remaining non-tested control questions, to which their performance would be compared. The order in which the passages were studied, the passages assigned to the different testing conditions, the order in which the tests were administered, and which 10-item question set was presented during the initial tests were counterbalanced across participants.

Results

Initial Test Performance

Participants in the 5-min and 48-hr delay conditions correctly answered 63% (*SD* = 18%) and 66% (*SD* = 19%) of the questions on the initial multiple-choice test, respectively. Participants in the 5-min and 48-hr delay conditions correctly answered 40% (*SD* = 21%) and 39% (*SD* = 18%) of the questions on the initial cued-recall test, respectively. As would be expected, initial test performance did not differ between the two retention-interval conditions.

Final Test Performance for Previously Tested Information

Final-test performance for previously tested information and comparable control information is presented in Figure 1. As indicated there, taking an initial test improved performance for those previously tested items on the later test, as compared to not taking an initial test, regardless of the type of initial test taken (cued-recall or multiple-choice) or the delay from initial test to the final test (5 min or 48 hr).

Interestingly, although performance appeared to be lower at a 48-hr delay than at a 5-min delay for questions initially presented in a multiple-choice test and for control questions (i.e., questions not previously tested), no forgetting appeared to occur for those items initially presented in a cued-recall test, suggesting a possible interaction. A 3 × 2 mixed-model ANOVA, however, did not reveal this apparent interaction between testing condition (cued-recall vs. multiple-choice vs. control) and delay (5 min vs. 48 hr), to be reliable, $F(2, 69) = 1.97$, $p > .05$. A significant main effect of testing condition, however, was obtained, $F(2, 69) = 43.73$, $\eta p^2 = 0.56$, $p < .01$. Most importantly, taking a multiple-choice test was better for retention of previously tested information than was taking a cued-recall test, $t(71) = 4.32$, $d = 0.51$, $p < .01$.

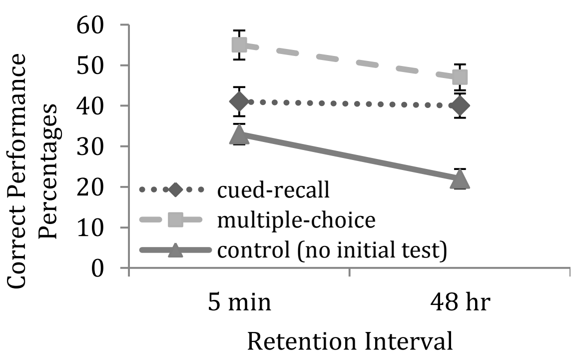

Figure 1: Correct performance percentages for previously tested information (and comparable non-tested control information) as a consequence of the testing condition (cued-recall, multiple-choice, or no-test control) and the retention interval between initial and final test (5 min or 48 hr). Error bars represent +/- 1 SE.

That performance for questions previously tested with a cued-recall test did not appear to differ when assessed at a 5-min delay versus a 48-hr delay is noteworthy, indicating the effectiveness of cued-recall tests for the retention of tested information and consistent with findings that items successfully retrieved at short delays tend to remain accessible at longer delays (e.g., Halamish & Bjork, 2011; Kornell, Bjork, & Garcia, 2011).

Final Test Performance for Non-tested Related Information

Results relevant to our primary objective of testing whether the finding that taking a multiple-choice test improves retention of non-tested information more than does taking a cued-recall test with a short retention interval between initial and final tests would replicate, and, if so, to assess whether such benefits would persist with a longer retention interval are shown in Figure 2. As can be seen there, information that was related to questions on an initial multiple-choice test appeared to be better recalled than was information related to questions on an initial cued-recall test or information from the non-tested control passage, both at a 5-min delay and at a 48-hr delay.

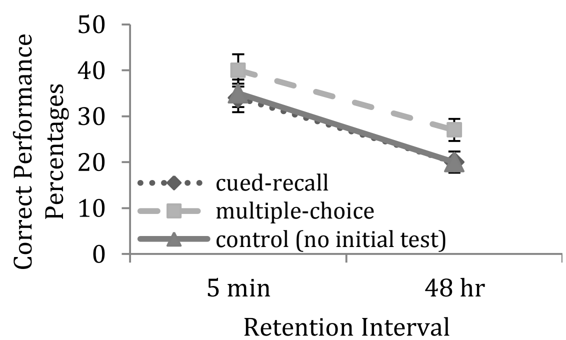

Figure 2: Correct performance percentages for non-tested related information (and comparable non-tested control information) as a consequence of the testing condition (cued-recall, multiple-choice, or no-test control) and the retention interval between the initial and final tests (5 min or 48 hours). Error bars represent +/- 1 SE.

Again, a 3 × 2 mixed-model ANOVA did not reveal an interaction between testing condition and delay, $F(2, 69) = 0.12$, $p > .05$, but did reveal a main effect of testing condition, $F(2, 69) = 5.04$, $\eta p^2 = 0.13$, $p < .01$. Overall, taking a multiple-choice test was better for the retention of related information than was either taking a cued-recall test or not taking a test (control condition), $t(71) = 2.72$, $d = 0.33$, $p < .01$ and $t(71) = 2.82$, $d = 0.33$, $p < .01$, respectively. Importantly, however, no difference in performance for the cued-recall condition versus the control condition was observed, indicating that although taking a cued-recall test did not impair the later recall of related information in the present study, neither did it help it compared to not taking a test at all. Taking a multiple-choice test did, however, and this benefit occurred regardless of the retention interval.

Intrusions of Incorrect Information

As previously mentioned, multiple-choice tests can result in misinformation effects (Roediger & Marsh, 2005; Marsh, Roediger, Bjork, & Bjork, 2007) thought to be a drawback to multiple choice testing as compared to more open-ended types of testing, and, indeed, participants in the present study made such intrusions on the final test in the multiple-choice condition. Because, however, all of the alternatives had appeared in the passages, participants also made such intrusions in the control and cued-recall conditions.

In the present study, there appeared to be more intrusions as a consequence of taking a multiple-choice test ($M = 24\%$, $SE = 3\%$) than as a consequence of taking a cued-recall test ($M = 12\%$, $SE = 2\%$) or having no initial test ($M = 14\%$, $SE = 2\%$) when the final test occurred at a 5-min delay. At the 48-hr delay, however, intrusions as a consequence of taking a multiple-choice test ($M = 17\%$, $SE = 2\%$) appeared to be reduced while intrusions occurring as a consequence of taking a cued-recall test ($M = 13\%$, $SE = 2\%$) or having no test ($M = 14\%$, $SE = 2\%$) did not. Looking specifically at the interaction between the two test types and the delay, a 2 × 2 mixed-model ANOVA revealed an interaction, $F(1, 70) = 4.24$, $p < .05$, with an independent samples *t*-test demonstrating that multiple-choice alterative intrusions were significantly lower at the 48-hr delay than at the 5-min delay, $t(70) = 2.14$, $d = 0.50$, $p < .05$. In considering these results, one should note that while correct performance in the control condition was lower at the 48-hr delay than at the 5-min delay, intrusion rates at the two delays were not different, suggesting that forgetting rates of correct and incorrect responses are not necessarily related. For the multiple-choice condition, however, both correct responses and intrusions were reduced at the 48-hr delay compared to the 5-min delay. Interestingly, in the multiple-choice condition, correct responses and intrusions were negatively correlated, $r(72) = -.30$, $p < .05$, suggesting that the more correct answers one recalled, the fewer intrusions one made, in both delay conditions.

Discussion

Taking a multiple-choice test shortly after study appears to have some particularly positive consequences for learning. Doing so not only increases retention of both tested and non-tested related information compared to taking a cued-recall test or no test, but such benefits also persist over a retention interval that is, by most standards, considered educationally realistic. In the present study, we also found that even misinformation effects, a frequently cited negative consequence of taking a multiple-choice test, become less of a problem at a longer retention interval. Thus, at a longer delay from initial to final test, multiple-choice testing still had the benefits seen at a shorter delay, but with less cost.

The present results can be contrasted with much past research demonstrating that cued-recall tests are better for retention than multiple-choice tests (e.g., Hamaker, 1986). Our finding of a greater benefit for multiple-choice testing than cued-recall testing might have occurred owing to the relatively low performance for cued-recall questions on our initial test: Although participants were unlikely to forget answers to questions that they answered correctly on the initial cued-recall test, the majority of responses on the initial test were incorrect and thus unlikely to be answered correctly later. In contrast, more of the multiple-choice questions were answered correctly on the initial test. Thus, although answering a given multiple-choice question might be less powerful for long-term retention than answering a given cued-recall question (Foos & Fisher, 1988), the greater number of correct responses on the initial multiple-choice test would result in more items having the potential to be recalled correctly later. It should be noted, however, that our use of highly competitive alternatives might also have resulted in our multiple-choice questions being more powerful for retention than those typically used in past research. Even so, perhaps providing feedback, a common practice in educational contexts, would lead to a reversal in our observed effect because feedback provides a second opportunity for the correct answers to incorrectly answered cued-recall questions to be learned (e.g., Kang, McDermott, & Roediger, 2007). Importantly, however, Little et al. (in press) also demonstrated that—although providing feedback did improve performance for information initially tested with a cued-recall test more than for information initially tested with a multiple-choice test, a reversal in final test performance was not realized. Furthermore, performance for related information was not affected as a consequence of whether or not feedback was provided.

Additionally, in the present study, we did not find evidence of retrieval-induced forgetting in the cued-recall condition, which is interesting, particularly at the short delay and given the competitive nature of our question pairs. Although retrieval-induced forgetting has been demonstrated with a variety of materials (see, e.g., Bjork, Bjork, & MacLeod, 2006; Storm, 2011), its occurrence with educational text materials is less consistent, with some studies observing retrieval-induced forgetting (e.g., Carroll et al., 2007; Little et al., 2011) and others not (e.g., Chan et al., 2006). Competition during the initial test, coherence of the to-be-learned materials, and the delay between initial and final tests have been pointed to as predictors of whether or not the effect will occur, but the exact specifications of the boundary conditions for the occurrence, or lack thereof, of retrieval-induced forgetting remain to be determined.

Concluding Comments

Multiple-choice tests are widely used in educational contexts and elsewhere, but their use—either as instruments of assessment or learning—is frequently disparaged. Although the present research does not directly speak to their potential value as tools of assessment, it does speak to their use as tools for learning. And, in that respect, as tools to reinforce knowledge, for the long as well as the short term, they seem quite useful when constructed with competitive alternatives and, perhaps, uniquely so with respect to the increased learning of competitive related as well as actually tested information.

Acknowledgments

A collaborative research grant from the James S. 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.

Anderson, R. C., & Biddle, W. B. (1975). On asking people questions about what they are reading. In G. H. Bower (Ed). *The Psychology of Learning and Motivation* (Vol.9). New York, NY, US: Academic Press.

Bjork, E. L., Bjork, R. A., & MacLeod, M. D. (2006). Types and consequences of forgetting: Intended and unintended. In L-G Nillson and N. Ohta (Eds.), *Memory and Society: Psychological Perspectives.* Psychology Press: Hove and New York.

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

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. K. (2009). When does retrieval induce forgetting and when does it induce faciliation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, *61*, 153-170.

Chan, J. C. K., 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.

Foos, P. W., & Fisher, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology, 80,* 179-183.

Frase, L. T. (1971). Effect of incentive variables and type of adjunct questions upon text learning. *Journal of Educational Psychology*, *62,* 371-375.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 801-812.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, *56*, 212-242.

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.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language,* 65,85–97.

Little, J. L., & Bjork, E. L. (2010). Multiple-choice testing can improve the retention of non-tested related information. In S. Ohisson & R. Catrabone (Eds.) *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1535-1540). Austin, TX: Cognitive Science Society.

Little, J. L., & Bjork, E. L. (2011). Pretesting with multiple-choice questions facilitates learning. In L. Carlson, C. Hőlscher & T. F. Shipley. (Eds.) *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 294-296). Austin, TX: Cognitive Science Society.

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (in press). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*.

Little, J. L., Storm, B. C., Bjork, E. L. (2011). The Costs and Benefits of Testing Text Materials. *Memory*, *19,* 346-359

Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, *14*, 194-199.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval induced forgetting. *Psychological Science, 18,* 29-34.

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

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

Rothkopf, E. Z. (1970). The concept of mathemagenic activities. *Review of Educational Research, 40,* 325-336.

Storm, B.C. (2011). Retrieval-induced forgetting and the resolution of competition. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in honor of Robert A. Bjork.* New York, NY: Psychology Press.