MEMORY, 2011, 19 (4), 346 359

Psychology Press
Taylor & Francis Group

# The costs and benefits of testing text materials

**Jeri L. Little, Benjamin C. Storm, and Elizabeth Ligon Bjork**

Department of Psychology, University of California, Los Angeles, CA, USA

Tests have been shown to improve the later recall of tested information, a result known as the testing effect. Tests, however, can also impair the later recall of related information, an effect known as retrieval induced forgetting. Although retrieval induced forgetting has been demonstrated using a wide variety of materials, recent work suggests that learning information in the context of a coherent text passage may afford protection from retrieval induced forgetting. In four experiments we explored the conditions under which retrieval induced forgetting does and does not occur with such materials. We found that two factors the coherence of the to be learned material and the competitiveness of retrieval practice are important in determining whether retrieval induced forgetting does or does not occur. Furthermore, even if retrieval induced forgetting does occur, having the opportunity to restudy the forgotten information can prevent that forgetting from persisting. Taken together, these findings provide greater understanding of the costs and benefits of testing text materials, with possible implications for the optimisation of testing as a tool for learning in educational contexts.

***Keywords:*** Testing effects; Retrieval-induced forgetting; Coherent text; Memory.

Although tests are primarily used in educational settings for assessment, they also provide opportunities for additional learning. When information is successfully retrieved from memory, its representation in memory is changed such that it becomes more recallable in the future and to a greater extent than would result from additional study (Bjork, 1975). Such benefits of testing on tested items have been well established (see, e.g., Carrier & Pashler, 1992; Landauer & Bjork, 1978; Roediger & Karpicke, 2006). Many questions remain, however, regarding the consequences of testing on related, but not initially tested items.

Many associative theories of memory (e.g., spreading activation, Collins & Loftus, 1975; Adaptive Control of Thought-Rational (ACT-R), Anderson, 1996; and Search of Associative Memory, Raaijmakers & Shiffrin, 1981) assume that when information is activated via retrieval, related information should also become activated. Recalling that toucans live in the rainforest, for example, might also bring to mind a picture of their colourful bills and knowledge of what they eat. Accordingly, in addition to enhancing the later recall of tested items (e.g., where toucans live), testing might also enhance the later recall of untested items (e.g., what toucans eat). Sometimes, however, the attempt to retrieve a target piece of information leads to the activation of other concepts or targets that have the potential to interfere with the retrieval of the desired information. When some information interferes with one's ability to recall other desired information, that interfering information needs to be selected against and consequently may be forgotten (see, e.g., Anderson, Bjork, & Bjork, 1994;

Address correspondence to: Jeri L. Little, Department of Psychology, University of California, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095, USA. E mail: jerilittle@ucla.edu

Benjamin C. Storm is now at Department of Psychology, University of Illinois at Chicago. The present research was supported in part by Grant 29192G from the James S. McDonnell Foundation. We thank Dr Robert Bjork, Dr Nate Kornell, and members of CogFog for helpful comments. Some of these results were presented at the 2007 Annual Convention of the American Psychological Association in San Francisco, CA.


http://www.psypress.com/memory DOI:10.1080/09658211.2011.569725

Anderson & Neely, 1996). That is, the retrieval of a subset of items can impair—rather than facilitate—the later recall of related non-retrieved items.

Anderson et al. (1994) demonstrated such impairment using a retrieval-practice paradigm consisting of three phases: study, retrieval practice, and final test. Participants first studied category–exemplar pairs (e.g., *Fruit–Banana*, *Tree–Elm*). Then, during the retrieval-practice phase, they practised retrieving half of the items from half of the categories in response to a series of category-plus-two-letter-stem retrieval cues (e.g., *Fruit–Ba* ). Following a 20-minute filled delay, a final test was given for items that had received retrieval practice (Rp+), for unpractised items from practised categories (Rp−), and for unpractised items from unpractised categories (Nrp), which served as baseline control items. On this final test, Rp+ items were better recalled than were Nrp items, replicating the typical testing effect. Rp− items, however, were less well recalled than were Nrp items—an effect referred to as *retrieval-induced forgetting*. Retrieval-induced forgetting has since been observed in a variety of other paradigms and for different types of material (e.g., Ciranni & Shimamura, 1999; Levy, McVeigh, Marful, & Anderson, 2007; Macrae & MacLeod, 1999; Radvansky, 1999; Saunders, Fernandes, & Kosnes, 2009; Shaw, Bjork, & Bjork, 2005), and the term, in general, refers to the impaired recall of some items resulting from the previous retrieval of related items.

Although shown to be a highly robust and general phenomenon, there are nonetheless conditions that appear to protect related unpractised information from suffering retrieval-induced forgetting. For example, Anderson and McCulloch (1999) found significantly less forgetting when participants were told to integrate study items—or when they did so spontaneously. Similarly, Anderson, Green, and McCulloch (2000) found that when practised and unpractised items were highly similar, retrieval-induced forgetting was not only eliminated, it was reversed (see also Garcia-Bajos, Migueles, & Anderson, 2009). Episodic or semantic integration may reduce competition during retrieval practice—and when retrieval practice is made non-competitive, very little if any retrieval-induced forgetting occurs (e.g., Anderson et al., 1994; Anderson, Green, et al., 2000; Shivde & Anderson, 2001; Storm, Bjork, & Bjork, 2007; see Storm, 2010, for a review of the role of competition in retrieval-induced forgetting). Additionally, integration has been shown to provide protection in other contexts where competition might lead to negative consequences (e.g., increases in errors or reaction time). For example, Smith, Adams, and Schorr (1978) demonstrated that *fan effects* (i.e., the finding that retrieval time and error rates increase when multiple disparate facts are learned about a certain concept; see, e.g., Anderson, 1974) that would occur given two sentences about one person (e.g., *Marty broke the bottle* and *Marty did not delay the trip*) were reduced when given an additional integrating fact that clarified a connection between the sentences (*Marty was chosen to christen the ship*). In short, if practised and unpractised information is integrated, unpractised information may interfere less with the retrieval of practised information, thus reducing the likelihood that it will suffer from retrieval-induced forgetting.

## RETRIEVAL-INDUCED FORGETTING AND EDUCATION

The occurrence of retrieval-induced forgetting suggests a possible negative consequence to test taking in educational settings: Namely, after an initial test of some items, students' recall of related items not appearing on that initial test may be impaired should they appear on a later more comprehensive test or exam. Although some evidence suggests that retrieval-induced forgetting can persist over a delay of 24 hours from retrieval-practice to final test (Migueles & Garcia-Bajos, 2007; Storm, Bjork, Bjork, & Nestojko, 2006), other evidence suggests that the impairment does not persist at these longer delays (e.g., MacLeod & Macrae, 2001; Chan, 2009). If retrieval-induced forgetting is a short-term effect, one might conclude that the effect is not a cause for concern in educational settings as teachers are unlikely to give two tests on the same topic in the same day. On the other hand, not all retrieval events in educational settings occur in the form of tests or quizzes that instructors may give their students days or weeks before administering a more comprehensive exam on the same subject matter. The use of flashcards and study guides or practice tests is common among students, and instructors often find students using such study aids even minutes before they administer a

comprehensive exam. Accordingly, when shortly before a comprehensive exam, students practise retrieving information using such devices (which most likely only sample a subset of the to-be-learned information), an inadvertent consequence may be an impaired ability to recall related, but initially untested, material on the actual exam.

Then again, as might be inferred from the previously mentioned findings of Anderson and colleagues (Anderson & McCulloch, 1999; Anderson, Green, et al., 2000; Garcia-Bajos et al., 2009) showing that integration of study materials either significantly reduces or eliminates the occurrence of retrieval-induced forgetting, unpractised information may be protected from such forgetting when the to-be-learned information is presented in a manner that encourages integration, as may be the case in some educational materials (e.g., coherent text passages). Text passages used in educational contexts often contain many facts, but such facts are often connected by sentences that elucidate the relationship between the facts (e.g., sentences that compare or contrast one idea with another, or clarify a causal or conceptual relationship between successive facts).

Despite the importance of discerning—within an educational context—the circumstances under which the testing of some information induces the forgetting of untested related information, very little research has examined this potential effect using educationally realistic materials. Moreover, the few efforts to address this issue have achieved mixed results (e.g., Camp & deBruin, 2008; Carroll, Campbell-Ratcliffe, Murnane, & Perfect, 2007; Chan, 2009; Chan, McDermott, & Roediger, 2006; Macrae & MacLeod, 1999).

Chan et al. (2006) demonstrated that the selective retrieval of a subset of information from text passages can not only protect against retrieval-induced forgetting, it can actually improve the later recall of related untested information. Utilising a scientific article about toucans, Chan et al. developed pairs of semantically related questions. For example, the questions "*What other bird species is the toucan related to?*" and "*Toucans sleep inside* " were semantically related because, as the text states, toucans cannot make their own *tree holes* and therefore must rely on other animals, like their relative the *woodpecker*, to make the tree holes in which they sleep. Participants initially read the article and were then assigned to one of three conditions: retrieval-practice (one question from each pair was tested), extra-study (one statement from each pair, constructed from a question and answer, was re-studied), or control (no retrieval practice or extra study was given). Following a 24-hour delay, participants in the retrieval-practice condition were more likely to recall the answers to the related questions that were not tested during retrieval practice than were participants in the other two conditions to answer their corresponding questions. These results thus suggest that under some conditions, retrieval of some information can induce facilitated retrieval of related information, rather than forgetting.

The fact that Chan et al. (2006) demonstrated retrieval-induced facilitation was both important and surprising because prior work employing educational materials had observed the typical pattern of retrieval-induced forgetting. Macrae and MacLeod (1999), for instance, asked participants to learn facts about two fictitious islands in preparation for a later final exam. After studying the facts about each island, which were similar to facts one might learn about different locals in a geography class, participants received retrieval practice for half of the facts about one of the islands. Then, on a later free-recall test, participants recalled significantly fewer of the unpractised facts about the practised island (Rp − items) than unpractised facts about the unpractised island (Nrp items). Thus Macrae and MacLeod found retrieval-induced forgetting even when participants—expecting a later comprehensive test—should have been motivated to remember all of the presented facts about both islands.

Perhaps significant for the finding of these disparate results, these two studies differed in numerous ways, including the nature of their study materials, the relationship between Rp+ and Rp − items, and the delay employed between retrieval practice and final test. First, with respect to study materials, Macrae and MacLeod (1999) had participants study information in the context of randomly ordered individual facts, whereas Chan et al. (2006) had participants study information in the context of a coherent text passage. In addition, Chan et al. designed their question pairs such that the items tested during the initial retrieval-practice phase (i.e., their Rp+ items) would be systematically related to items that were not initially tested, but that would then appear on the final test (i.e., their Rp − items). Although there are a variety of ways in which two questions might be systematically related, the pairs used by Chan et al. were created based on conceptual relatedness and close temporal proximity in the

passage. As a consequence, sometimes information contained in one question in the pair could be used to aid in the retrieval of the answer to the other question in the pair. Thus the question pairs had a facilitative relationship, such that retrieving the answer to one question in a given pair might encourage the spontaneous recall of information related to the other question in the pair. Chan et al. also determined that pairs were related using quantitative estimates including latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998). In contrast, Macrae and MacLeod did not design the questions they used during retrieval-practice (or the initial test) in this manner. Facts were about separate topics and there was no systematic manipulation of temporal proximity during the study phase. Lastly, with respect to retention intervals, whereas Chan et al. (2006) administered their final test 24 hours after the initial retrieval-practice phase, Macrae and MacLeod (1999) administered their final test after only a 5-minute delay. As previously discussed, delay between retrieval practice and final test may be important in determining whether retrieval-induced forgetting occurs (Macrae & MacLeod, 2001). Possibly, then, any of these differences—use of coherent text rather than isolated facts as the to-be-learned materials; pairs of questions used in the initial and final tests that were specifically designed to be facilitative; or longer retention intervals between retrieval practice and final test—could have contributed to the facilitation of initially untested related information as observed by Chan et al. (2006) rather than its forgetting as observed by Macrae and MacLeod (1999).

Relevant to this issue is research by Carroll et al. (2007) in which they explicitly investigated whether the presentation of to-be-learned material as coherent text might afford some protection from the occurrence of retrieval-induced forgetting and found that doing so did not render such material ''immune from the detrimental effects of retrieval practice'' (p. 592). Specifically, these investigators presented the to-be-learned information either in the form of a coherent text passage or as disordered sentences and then tested both practised and unpractised information via essay, short-answer, or multiple-choice tests after a brief delay, and they found no evidence of text coherence rendering protection from retrieval-induced forgetting, with retrieval-induced forgetting occurring on both the essay and short-answer tests for both ordered and disordered sentences (although not for the multiple-choice tests). In another study, however, they did find expertise with the particular subject matter of the text to provide some protection in that their expert participants suffered less retrieval-induced forgetting than did their novice participants, with amount of forgetting diminishing for both groups over a 24-hour period between retrieval practice and final test. For neither group, however, did they observe any sign of facilitation for previously untested information, even at the 24-hour retention interval and even though materials were originally studied in the form of coherent text.

In contrast to these findings of Carroll et al. (2007) indicating that coherence does not protect against retrieval-induced forgetting, recent work by Chan (2009), in which coherence of the to-be-learned material was also manipulated, produced a different pattern of results. More specifically, Chan presented to-be-learned information in either a high integration condition (i.e., as a coherently ordered passage) or in a low integration condition (i.e., the same passage with its paragraphs in the same order but the sentences within each paragraph randomly ordered) and, in addition, manipulated the length of the delay between retrieval practice and final test (20 minutes vs. 24 hours). Similar to the results of Carroll et al., Chan found retrieval-induced forgetting in the low integration condition at the 20-minute delay; however, unlike Carroll et al., he did not find forgetting in the high integration condition at the 20-minute delay. Moreover, at the 24-hour delay, no retrieval-induced forgetting was observed in the low-integration condition, and retrieval-induced facilitation was observed in the high-integration condition, with this latter finding thus replicating Chan et al. (2006). Finally, investigating the potential effect of initial testing at three different delays from the retrieval-practice phase, Chan (2010) again found facilitation at a 24-hour delay, which still persisted at a 7-day delay; but at a 20-minute delay he found neither facilitation nor forgetting.

In summary, then, these studies have produced a conflicting pattern of results with respect to the effects of prior testing when the to-be-learned material is presented in the form of coherent text. At long delays between retrieval practice and final test, Chan (2009, 2010) has observed enhanced recall of related untested information; whereas Carroll et al. (2007) have not, although Carroll

et al. did observe diminished retrieval-induced forgetting at a 24-hour delay. At shorter delays between retrieval practice and final test, the results are also contradictory, with the potential integration afforded by coherent text sometimes seeming to protect against retrieval-induced forgetting (Chan, 2009, 2010) and sometimes not (Carroll et al., 2007). Thus a primary goal of the present research was to explore when integration in the form of text coherence serves as a boundary condition for retrieval-induced forgetting in situations where short delays are used and to identify, more generally, the conditions under which retrieval-induced forgetting does and does not occur with text material. Additionally, we explored whether—even when retrieval-induced forgetting does occur despite the use of coherent text materials—there are ways to reverse such forgetting.

## EXPERIMENT 1

One possible contributor to the inconsistent results found by Chan (2009, 2010) and Carroll et al. (2007) in their studies manipulating coherence of the to-be-learned material may have been the use by Chan of questions in the initial test or retrieval practice (i.e., the Rp+ items) that were each paired with a highly associated, but non-competitive, question that was then asked in the final test (i.e., as Rp− items). That is, Chan specifically designed his pairs of Rp+ and Rp− items to be facilitative (i.e., so that the recall of one item would be likely to encourage the spontaneous recall of information related to the paired item owing to a conceptual relationships between the items or close temporal proximity in the passage) and, perhaps, without the use of such items, Chan—like Carroll et al. (2007)—would also have observed retrieval-induced forgetting when information was presented as coherent text.

We examined this possibility in Experiment 1 by using pairs of Rp+ and Rp− items that—although created from text materials highly similar to those used by Chan (2009, 2010)—were not specifically designed to be facilitative. Instead, our pairs of Rp+ and Rp− items were randomly formed for counterbalancing purposes only, and the items within a pair were only related in that they were questions based on facts presented in the same passage about a particular topic. If, using these non-facilitative Rp+ and Rp− items, we find retrieval-induced forgetting to occur in the final test, consistent with the results of Carroll et al. (2007), then it seems likely that the facilitative relationship between Rp+ and Rp− items during retrieval practice in the Chan studies was responsible for his not observing retrieval-induced forgetting with his coherent text materials.

### Method

*Participants and design.* A total of 72 UCLA undergraduates served as participants for credit in their introductory psychology course. Item type (Rp+, Rp−, or Nrp) was manipulated within participants; that is, all participants served in both retrieval-practice and control conditions.

*Materials.* Two 800-word passages were constructed so that, for each participant, one could serve as the practised passage (i.e., the passage about which a subset of questions were asked after its presentation and before the final test) and one could serve as the control or unpractised passage (i.e., the passage about which no questions were asked until the final test). One passage was about toucans (a truncated version of the passage used by Chan et al., 2006) and one was about the planet Mercury. A set of 24 fill-in-the-blank questions was developed for each passage. These questions were not constructed to be facilitative and, after construction, were simply randomly divided into two subsets that served equally often as Rp+ and Rp− items when that passage was practised and as the corresponding Nrp items when that passage was not practised. For most questions, 1–2 words or a 2–3-digit number provided a sufficient answer (e.g., for the question: "*Due to the recent reclassification of Pluto from status as a planet to that of a , Mercury has now regained the 'smallest planet' title*" the correct answer was: "*dwarf planet*").

*Procedure.* All participants read both passages, one sentence at a time, on a computer screen. The order of the passages was counterbalanced such that half of the participants read the Toucan passage first and the other half read the Mercury passage first. Each passage took approximately 6 minutes to present, and each sentence was presented for a time appropriate to its length. After the initial study phase participants worked on an arithmetic problem for 30 seconds and then engaged in a retrieval practice phase consisting of 12 fill-in-the-blank questions for one of the

passages. After all questions had been tested once, the same questions were tested again, but in a different random order. Each question appeared for 12 seconds, and participants had the full 12 seconds to type their answer on the keyboard.

After a 15-minute non-verbal distraction task (playing Tetris), participants received a final fill-in-the-blank test for all of the questions pertaining to both passages. The questions appeared on a computer screen and participants were given 15 seconds to answer each question. Thus, in the final test for a given participant, there were 12 questions that had been tested previously (i.e., Rp+ items), 12 questions from the same passage that had not been tested previously (i.e., Rp− items), and 24 questions about the other passage for which no questions had been tested previously (i.e., Nrp items). To control for output interference, Rp− items were always tested before Rp+ items.

## Results and discussion

In the analyses reported below for Experiments 1 and 2 we compare recall performance for Rp− items and Rp+ items to different Nrp control items. As Rp− items were always tested first on the final test, their recall is compared to recall performance for Nrp items also tested first on the final test, which we refer to as Nrpa items. Recall performance on Rp+ items, which were always tested second on the final test, is compared to that for Nrp items also tested second on the final test, which we refer to as Nrpb items. Additionally, the Nrp items were counterbalanced such that they served equally often as control items for Rp− items (i.e., as Nrpa items) and as control items for Rp+ items (i.e., as Nrpb items).

*Retrieval-practice performance.* Participants correctly answered 47% ($SD = 19\%$) of the questions during retrieval practice, and their performance was significantly better on the second round of retrieval practice ($M = 48\%$, $SE = 2\%$) than on the first round ($M = 46\%$, $SE = 2\%$), $t(71) = 2.57$, $p < .05$.

*Final-test performance.* A significant testing effect emerged such that Rp+ items ($M = 49\%$, $SE = 2\%$) were recalled significantly better than comparable Nrpb items ($M = 41\%$, $SE = 2\%$), $t(71) = 3.33$, $p < .01$. Of primary interest, however, was the fate of the Rp− items: Namely, would their recall be impaired—that is, show retrieval-induced forgetting—under the present conditions? A planned-comparison $t$ test revealed this not to be the case, with recall of Rp− items not differing significantly from that of their comparable Nrp items ($M = 42\%$, $SE = 2\%$ and $M = 43\%$, $SE = 2\%$ for Rp− and Nrpa items, respectively), $t(71) = 0.57$, $p > .05$.

Thus, even without the use of Rp+ and Rp− questions specifically designed to be facilitative, giving participants an initial test on a subset of questions about the text passage failed to produce retrieval-induced forgetting of the untested questions for that passage. That is, despite our employing conditions in which retrieval-induced forgetting is typically observed, no evidence of retrieval-induced forgetting was seen in the present Experiment 1. One possibility is that the coherence of the text led to the spontaneous integration of to-be-learned facts. When reading a given sentence, participants likely considered that sentence in light of the previous sentence—and to the extent that the previous sentence provided a transition from earlier facts, those separate facts may have been integrated. These results are thus consistent with the notion that the coherence of text can, at least sometimes, provide a learning context that protects untested material from retrieval-induced forgetting and, furthermore, can do so even when the questions serving as Rp+ and Rp− items are not designed to be facilitative.

As indicated by the study of Chan (2009), this protection most likely arises owing to the integration of information that is promoted by the coherence of the text in which it is presented. Indeed, when Chan presented the same information in a less-coherent form, retrieval-induced forgetting was observed at the short retention interval. However, because this differential effect of coherently ordered information versus disordered information was not also found by Carroll et al. (2007), we examined the role of coherency again in the present Experiment 2. Specifically, we presented the same text passage as in Experiment 1, but—rather than presenting the sentences in a coherent order—we presented the sentences in a random order, similar to the procedure used by Carroll et al. and Chan. By so reducing the coherence of the text, we expected to prevent participants from spontaneously inter-relating information while they read, thus making

the information vulnerable to retrieval-induced forgetting.

## EXPERIMENT 2

### Method

*Participants.* A total of 72 UCLA undergraduates served as participants for course credit. As in Experiment 1, item type (Rp+, Rp−, or Nrp) was manipulated within participants; that is, all participants served in both retrieval-practice and control conditions.

*Materials and procedure.* The materials and procedure were the same as those used in Experiment 1 except that, during the reading of the materials, the sentences comprising each passage were presented in a random order. Each sentence, however, was understandable on its own. A Latent Semantic Analysis (LSA; see Landauer et al., 1998) for sentence-to-sentence comparisons, in which a rating of −1 indicates no coherence and +1 indicates perfect coherence, confirmed that the text passages ($M = 0.35$, $SE = 0.03$) used in Experiment 1 were more coherent than were the randomly ordered presentations ($M = 0.25$, $SE = 0.02$) used in Experiment 2, $t(140) = 2.75$, $p < .01$.

### Results and discussion

*Retrieval-practice performance.* Participants answered 45% ($SD = 20\%$) of the retrieval practice questions correctly. As in Experiment 1, performance was significantly better in the second round of retrieval practice ($M = 47\%$, $SE = 2\%$) than it was in the first ($M = 43\%$, $SE = 2\%$), $t(71) = 4.10$, $p < .001$.

*Final-test performance.* We once again replicated the testing effect, finding the Rp+ items ($M = 48\%$, $SE = 2\%$) to be recalled significantly better than the Nrpb control items ($M = 42\%$, $SE = 2\%$), $t(71) = 2.67$, $p < .01$. Additionally, with respect to the question of primary interest, the untested Rp− items suffered from retrieval-induced forgetting, with recall performance for Rp− items ($M = 41\%$, $SE = 2\%$) being significantly worse than that for the Nrpa control items ($M = 46\%$, $SE = 2\%$), $t(71) = 2.30$, $p < .05$. Thus, as in the Chan (2009) study, when the sentences of the text were no longer presented in a coherent order, Rp− items were no longer immune to retrieval-induced forgetting.

Taken together, the results of Experiments 1 and 2 provide a replication of two recent and independent lines of research (Chan, 2009, and Camp & deBruin, 2008) suggesting that coherent text can afford protection from retrieval-induced forgetting. However, these results contrast with those of Carroll et al. (2007, Experiment 2), which did not show such protection for coherent text.

One explanation of why presentation of to-be-learned materials in the form of coherent text sometimes does and sometimes does not protect untested related material from retrieval-induced forgetting can be found in the inhibitory account of retrieval-induced forgetting (see Anderson, 2003; Anderson et al., 1994; Bjork, Bjork, & Anderson, 1998). According to this account, retrieval-induced forgetting occurs as a result of inhibitory processes that act during retrieval practice to resolve competition from untested related items. Thus what is critical for the observance of retrieval-induced forgetting, or the lack thereof, is whether there is competition during retrieval practice, and several lines of research have clearly demonstrated this competition-dependent nature of retrieval-induced forgetting (e.g., Anderson et al., 1994; Anderson, Bjork, & Bjork, 2000; Shivde & Anderson, 2001; Storm et al., 2007; for a review see Storm, 2010). The coherence of the passages in Experiment 1 led to the protection of untested related information from retrieval-induced forgetting, presumably owing to the integration of the information afforded by such text. When sentences were randomised in Experiment 2, however, integration across the sentences could not easily occur, and protection from retrieval-induced forgetting was not realised. This is not to say, however, that the use of coherent text should always result in protection from retrieval-induced forgetting. For example, although a coherent text passage may afford the integration of information within that passage, information across multiple passages may not be integrated together even though each passage is presented in the form of coherent text. To the extent that competitive information (e.g., multiple targets that have an association with a given cue) occurs across separate coherent texts, practising information from one text should lead to the forgetting of information in another text. That is, information contained in one text passage should be vulnerable to forgetting when related information from another text passage is

practised, provided that there are competitive relationships among the information presented in the separate passages.

We explored this possibility in Experiment 3 by constructing individual passages that presented three categories of information (geography, climate, and people) about each of six different regions of the world, similar to the type of information that students would be asked to learn in a geography class. Thus, within each passage, the same three cues (i.e., geography, climate, and people) occurred, but across passages, different target information was associated with each of these cues (e.g., for geography, vast deserts in Africa; fjords and glaciers in Norway; freezing tundra in Canada). Then, during retrieval practice, we asked participants to retrieve these three types of information from a subset of these regions, analogous to their being given a test on some, but not all, of the regions studied in a geography class. We expected that having participants selectively retrieve information about a subset of the regions would cause forgetting of comparable information about the untested regions despite all of the studied information having been presented in the context of coherent text. More specifically, we believed that the nature of the materials, although promoting the integration of information within a given region, would not engender the integration of similar, but competitive, material across regions. That is, although participants might be led to integrate different types of information about a particular region (e.g., the geography, people, and climate of Canada), they should be less likely to integrate similar information across different regions (e.g., the geography of Canada, Norway, and Africa). As such, attempting to retrieve information about one region would be likely to activate competing information about another region, making retrieval practice competitive and requiring participants to avoid retrieving information about other regions in order to retrieve information about the tested region—just the type of conditions under which retrieval-induced forgetting is typically observed.

Finally, in order to approximate typical educational conditions more closely, we made the retrieval practice task more open-ended than is usually the case in studies of retrieval-induced forgetting. Rather than guiding participants to retrieve an exemplar or a missing word in a sentence, participants in Experiment 3 were asked to provide essay-type answers. Specifically, participants were asked to describe the geography, climate, and people of a specific subset of the regions. To our knowledge, the present study is the first to examine retrieval-induced forgetting using such an open-ended form of retrieval practice.

## EXPERIMENT 3

### Method

*Participants and design.* A total of 16 UCLA undergraduates served as participants for credit in their introductory psychology course. Practice condition (practice vs. no-practice or control) was manipulated between subjects. Item type (Rp+ or Rp−) was manipulated within participants in the practice condition.

*Materials.* Six passages, with an average length of 206 words, were constructed about various regions of the world (specifically, Norway, Australia, Siberia, Canada, Africa, Greenland), with each passage including three separate paragraphs about the geography, climate, and people of that region, respectively. Latent semantic analysis for sentence-to-sentence comparisons confirmed the average coherence for each of the 18 paragraphs ($M = 0.36$, $SD = 0.15$) to be comparable to that of the text materials used in Experiment 1. Furthermore, each paragraph contained four critical facts selected for the purpose of scoring recall performance; thus, 12 critical facts were scored per passage.

*Procedure.* Participants were given a six-page packet with the information about each of the to-be-learned regions appearing on a separate sheet of paper. They were given a total of 10 minutes to study all six passages and could do so in any order. After this study phase a randomly selected half of the participants, who were assigned to the practice condition, were given two successive and identical recall tests, each lasting 6 minutes, during which they were to recall everything that they could remember about the geography, climate, and people of three of the six studied regions. The regions were randomly divided into two groups of three (which turned out to be Australia, Greenland, and Norway in one group and Siberia, Africa, and Canada in the other), and Siberia, Africa, and Canada were always the practised regions in the practice condition.

The remaining participants, who were assigned to the no-practice (Nrp) condition, were given a non-verbal distractor task for this 12-minute duration.

After a subsequent 15-minute delay, also filled with a non-verbal distractor task, participants in the practice condition were given 6 minutes to complete a recall test for all three unpractised regions (previously practised regions were not tested on the final test). Participants in the no-practice condition were also given 6 minutes to complete a recall test for the three passages corresponding to those on which the participants in the practice condition had been tested. Specifically, all participants were provided with three sheets of paper, each of which had the name of one region and the words *geography*, *climate*, and *people* and were asked to recall as much information about each of these categories for each of the three specified regions as possible.

## Results and discussion

*Retrieval-practice performance.* Of the 36 possible critical facts that could be recalled (12 from each of the three regions practised), participants in the practice condition successfully retrieved an average of 5.50 ($SE = 1.25$) and 5.00 ($SE = 1.21$) critical facts during the first and second blocks of retrieval practice, respectively.

*Final test performance.* In Experiment 3 neither Rp+ nor Nrpb items were tested on the final test; thus, we have used only the term Nrp to refer to control items. Recall performance on the final test as a function of practice condition was analysed using an independent-samples $t$ test. Significantly fewer facts about unpractised regions were recalled when other regions had been practised (Rp− items: $M = 3.00$, $SE = 0.63$) than when other regions had not been practised (Nrp items: $M = 5.50$, $SE = 0.73$), $t(14) = 2.59$, $p < .05$.

These results thus demonstrate that retrieval-induced forgetting can occur even when the to-be-learned information is presented in the context of coherent text, provided retrieval practice is competitive—that is, that the testing of some of the information results in competition from related, but untested information. Thus, even if text coherence can afford protection from retrieval-induced forgetting, it clearly does not always do so. Moreover, to the degree that the present materials are fairly typical of those used in the classroom, it seems likely that untested information might frequently suffer from retrieval-induced forgetting in educational settings.

However, the possibility of this impairment as a consequence of selective testing may not be as distressing as it first seems, because—even if related untested information does suffer from retrieval-induced forgetting at one point in time—its accessibility in memory may not necessarily be permanently impaired (e.g., MacLeod & Macrae, 2001). Moreover, evidence exists that forgetting can serve as an effective enabler of future learning (see, e.g., Bjork, 1994; Bjork & Bjork, 1992). It is possible, therefore, that items suffering from retrieval-induced forgetting could benefit more from subsequent relearning opportunities than items not suffering from retrieval-induced forgetting. Indeed, Storm, Bjork, and Bjork (2008) have demonstrated such a phenomenon using category–exemplar pairs. In their study not only was the effect of retrieval-induced forgetting eliminated following relearning, it was also reversed.

The goal of the present Experiment 4 was to see if the results of Storm et al. (2008) could be replicated using more educationally realistic materials. If so, such a finding would provide a fairly straightforward procedure for reversing the negative effects, while maintaining the positive effects, of selective testing. Employing the same materials as those used in Experiment 3 we provided some participants the opportunity to relearn information about the untested regions between retrieval practice and final test. We predicted that the forgetting observed in Experiment 3 would be eliminated following relearning and, possibly, that information initially forgotten might, as a consequence of relearning, become even more recallable than information that was not initially forgotten.

# EXPERIMENT 4

## Method

*Participants and design.* A total of 76 UCLA undergraduates served as participants for credit in their introductory psychology course. Practice condition (practice vs. no-practice or control) and relearning condition (relearning vs. no-relearning) were manipulated between participants. Item type (Rp+ or Rp−) was manipulated

within participants for those in the practice condition.

*Materials and procedure.* The materials and procedure were the same as those used in Experiment 3, with the exception that Experiment 4 incorporated a relearning phase for half of the participants. Following the non-verbal distractor task, a random half of the participants in both the practice and no-practice conditions were given a total of 4 minutes to restudy the passages about the three regions that had not received retrieval practice (i.e., the Rp − passages for the practice participants and the corresponding Nrp passages for the no-practice participants). The remaining half of the participants from each condition continued to work on the distractor task during this additional time. All participants were then given the same final recall test as was administered in Experiment 3. That is, participants were asked to recall as much information as they could about the three regions that had not been initially practised by the participants in the practice condition (i.e., the Rp − items), which at this point would have been restudied by a random half of both the practice and the no-practice participants and not restudied by the remaining half of the participants in these conditions.

## Results and discussion

*Retrieval-practice performance.* Of the 36 possible critical facts that could be recalled (12 from each of the three regions practised), participants in the practice condition successfully retrieved a total of 7.37 ($SE = 0.54$) and 7.16 ($SE = 0.55$) during the first and second blocks of retrieval practice, respectively.

*Final-test performance.* As in Experiment 3, neither Rp+ nor Nrpb items were tested on the final test in Experiment 4. Thus we again use only the term Nrp to refer to control items in the analyses reported below. Correct final recall scores for Rp − and Nrp items in the relearning and no-relearning conditions are shown in Figure 1. As in Experiment 3, recall performance on the final test was first analysed using an independent-samples $t$ test for participants who did not relearn the unpractised information. As indicated by the difference in the two left bars shown in Figure 1 and replicating the results of Experiment 3, significantly fewer facts about unpractised regions were recalled when other

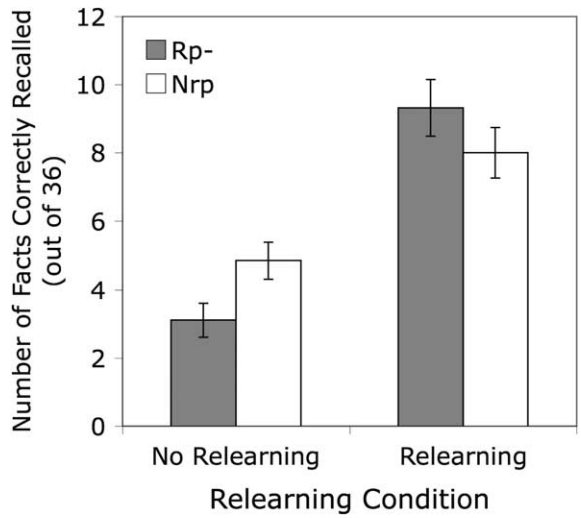


**Figure 1**. Mean number of facts correctly recalled on the final test as a function of item type (Rp or Nrp) and relearning condition (relearning or no relearning) in Experiment 4. Maximum number of items that could be recalled was 36. Error bars show mean ± 1.0 *SE*.

regions had been practised (Rp − items: $M = 3.11$, $SE = 0.50$) than when other regions had not been practised (Nrp items: $M = 4.54$, $SE = 0.53$), $t(36) = 2.39$, $p < .05$. That is, as in Experiment 3, we observed significant retrieval-induced forgetting.

Next we conducted the same analysis for participants who did relearn the unpractised information prior to taking the final test. Although not a significant difference, recall performance under these conditions, shown in the two right bars of Figure 1, was numerically better when other regions had been practised (Rp − items: $M = 9.32$, $SE = 0.82$) than when other regions had not been practised (Nrp items: $M = 8.00$, $SE = 0.74$), $t(36) = 1.18$, $p > .05$. Importantly, however, A $2 \times 2$ ANOVA indicated that the interaction between relearning and practice conditions was significant, $F(1, 72) = 5.27$, $p < .05$.

Experiments 3 and 4 demonstrated that selective testing can impair the later recall of related untested information—even when the to-be-learned information has been presented in the context of coherent text. As demonstrated by Experiment 4, however, such impairment can be eliminated if learners are re-exposed to the untested information prior to the final test. Accordingly, as long as learners have the opportunity to restudy related untested material, retrieval-induced forgetting should not persist. And indeed, as indicated by the numerically superior recall of the Rp − items versus the Nrp

items after restudy, as well as the findings of Storm et al. (2008), an opportunity for restudy may result in greater learning for items temporarily forgotten than for items not so forgotten.

## GENERAL DISCUSSION

Although retrieval-induced forgetting has been observed in a variety of learning contexts and with a variety of materials, ranging from category–exemplar pairs to materials in contexts that more closely resemble real-world situations, few studies have investigated its possible occurrence with the type of text materials learned in the classroom, where selective testing is perhaps most prevalent. The present research adds to this somewhat sparse literature by further exploring those conditions under which selective testing of information learned via the reading of text leads to retrieval-induced forgetting and when it does not.

Importantly, the results of Experiment 1 demonstrated that the apparent ability of text to protect related untested information from retrieval-induced forgetting does not depend on the use of tested and untested questions that are specifically designed to be facilitative. Rather, such protection appears to come from the opportunity afforded to learners for integrating information when it is presented in the form of coherent text, a hypothesis consistent with the findings across Experiments 1 and 2. Like Chan (2009) we failed to find evidence of retrieval-induced forgetting when items were presented in the context of coherently ordered text (Experiment 1), but did find evidence of forgetting when that same information was presented in a disordered manner (Experiment 2). Interestingly, other work has shown that—in some circumstances—a less-coherent text might lead to better learning than a more-coherent text because the lack of integration would necessitate more attention and active engagement (McNamara, Kintsch, Songer, & Kintsch, 1996). This finding might raise the question as to why our participants did not try to integrate the random sentences in order to obtain a better understanding of the to-be-learned-information. First, aspects of our procedure might have made it especially difficult for participants to integrate facts in the condition with randomly ordered sentences. In the present Experiments 1 and 2 sentences were shown one at a time on a computer screen, for a limited amount of time. Although there may be some advantage to processing disordered text when given ample time to study and perhaps the ability to revisit sentences (with such a presentation of materials functioning as a type of *desirable difficulty*; Bjork, 1994)—when time is limited and individual sentences are presented only once, processing disordered text may not afford such an advantage. Moreover, McNamara et al. found that coherence of text interacts with background knowledge such that participants with high background knowledge benefit from a less-coherent text, but participants with low background knowledge benefit from a more-coherent text. Although we did not have a measure of the background knowledge of our participants, we imagine that they would be more like the low-knowledge participants in the McNamara et al. study in terms of their level of prior knowledge about the subject matters of our materials in Experiments 1 and 2.

Additionally, in Experiments 3 and 4, we found that—although the learning of information via coherent text can sometimes afford protection from retrieval-induced forgetting—such is not always the case. More specifically, retrieval-induced forgetting can occur for information presented in the context of coherent text when the nature of that material fails to promote integration of tested and untested information and, thus, does not reduce competition between such information during retrieval practice—conditions that we believe were present in Experiments 3 and 4.

The present research thus offers a possible explanation of why the results obtained across the several recent studies investigating the role of integration in protecting information from retrieval-induced forgetting (Camp & deBruin, 2008; Carroll et al., 2007; Chan, 2009, 2010) have produced inconsistent results. Previous research on retrieval-induced forgetting has generally found that forgetting is only observed when there is a competitive relationship between tested and untested information (for a review see Storm, 2010). Such dynamics, we believe, explain our finding of retrieval-induced forgetting in the present Experiments 3 and 4. Similarly, they could also explain why retrieval-induced forgetting was observed by Carroll et al. (2007) but not by Chan (2009, 2010). Whereas the materials used by Chan were designed to be facilitative, the materials used by Carroll et al. (2007) appear, upon our inspection, to have been more likely to elicit the

type of competition necessary for retrieval-induced forgetting. For example, based on a passage they used concerning a patient with schizophrenia, their participants could be asked the following questions: *How could David's conversational speech be described?* or *David's tendency to jump from topic to topic and talk illogically is known as* . In this case these questions would provide similar cues and their targeted answers (i.e., *highly disorganised* and *associative splitting*, respectively) are likely to compete with one another because they could each serve as a plausible (although incorrect) answer to the other question. Thus we posit that if multiple plausible targets are contained within a coherent text passage, asking one question on an initial test (for which one target is the answer) and a similar, but different question on a later test (for which another target is the answer), may well lead to retrieval-induced forgetting. Individuals likely to be victims of this type of competition would be non-experts unfamiliar with these terms, and indeed such participants showed greater forgetting in Carroll et al.'s experiment.

Because our studies utilised short delays from retrieval practice to final test, the application of our findings to some educational situations (e.g., how a teacher should construct a quiz in preparation for a later exam; how students should test themselves the night before the exam) may be limited. As stated earlier, it is uncertain whether retrieval-induced forgetting persists over longer delays. Thus our results can only speak directly to situations in which retrieval of a subset of information occurs shortly before a more comprehensive exam. On the other hand, to the degree that the type of materials and retrieval practice employed in the present Experiments 3 and 4 are similar to how a student might self-test immediately before an exam (e.g., using flashcards or practice questions), our results suggest that there might be considerable potential for the occurrence of retrieval-induced forgetting to be a problem for students. On the bright side, however, Experiment 4 demonstrated that—even with text materials that do not lead to the type of integration that protects information from retrieval-induced forgetting—such forgetting can be eliminated when participants are given the opportunity to restudy the initially untested information prior to a final test. Although presenting the forgotten information (i.e., the Rp − items) for restudy in Experiment 4 did not result in their being recalled significantly better than corresponding items that had not suffered retrieval-induced forgetting (i.e., the control Nrp items), as had been found by Storm et al. (2008) with category–exemplar pairs, their impairment was eliminated and their recall was numerically greater than that of the information that had never been forgotten.

Thus, even if instructors inadvertently create retrieval-induced forgetting of related untested materials via selective testing (or if students create such consequences for themselves), the forgetting of such items can be eliminated if students have the opportunity to restudy them. Furthermore, given the numerically superior recall of Rp − items versus their Nrp controls in Experiment 4 as well as the previous results of Storm et al. (2008), restudy of such items may have the potential to result in the enhanced learning of such items. In other words, under the proper conditions, instructors' use of selective testing could lead not only to the positive benefits of testing, but also to the enhanced learning of the initially forgotten untested information—thus, a win-win situation for both tested and untested information.

## CONCLUDING COMMENTS

The results obtained across the present four experiments have both important theoretical and applied implications. From a theoretical standpoint, they more clearly delineate the conditions under which selective testing of material presented as coherent text results in retrieval-induced forgetting, pointing to the roles of integration and competition during retrieval practice as critical determiners of whether it will or will not occur. When the integration afforded by coherent text promotes the integration of tested and untested material, such that the latter does not compete during retrieval practice of the former, then retrieval-induced forgetting of the untested material can be prevented. If, however, the integration afforded by the textual context does not promote such integration—allowing the related untested information to compete during the retrieval practice of the tested material—then the untested material will continue to suffer from retrieval-induced forgetting.

From an applied perspective, the present results have important implications for the use of testing in educational practice and particularly for some of students' studying activities (e.g.,

self-testing shortly before an exam). The picture, however, is somewhat complicated. For example, our results demonstrate that simply advising teachers to imbed the to-be-learned information in coherent text to avoid inadvertently creating effects of retrieval-induced forgetting as a consequence of selective testing would not be sufficient. Rather, one needs to attend to the type of integration that the particular text affords and then try to ensure that it promotes integration across information that would otherwise compete during selective testing of that information. Similarly, students should integrate materials to be learned before giving themselves self-administered tests, such as they might do with flashcards that test some, but not all, of the material, particularly if they test themselves immediately before exams.

Additionally, the present results show the critical importance of including review sessions before final exams, whether given by the teacher or self-administered by the student. Even if prior selective tests and quizzes have caused the forgetting of related untested material, as long as that material is reviewed prior to a final exam, then that forgetting can be eliminated and, potentially, turned into facilitation. Hence, when coupled with proper review strategies, use of selective testing—either by the teacher or the student—can result in the benefits of testing without engendering the costs of retrieval-induced forgetting.

Manuscript received 29 September 2010
Manuscript accepted 16 February 2011

## REFERENCES

Anderson, J. R. (1974). Retrieval of propositional information from long term memory. *Cognitive Psychology*, *6*, 451 474.

Anderson, J. R. (1996). ACT A simple theory of complex cognition. *American Psychologist*, *51*, 355 365.

Anderson, M. C. (2003). Rethinking interference the ory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415 445.

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

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval induced forgetting: Evidence for a recall specific mechanism. *Psychonomic Bulletin & Review*, *7*, 522 530.

Anderson, M. C., Green, C., & McCulloch, K. C. (2000). Similarity and inhibition in long term mem ory: Evidence for a two factor model. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1141 1159.

Anderson, M. C., & McCulloch, K. C. (1999). Integra tion as a general boundary condition on retrieval induced forgetting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 608 629.

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of percep tion and cognition* (pp. 237 313). San Diego, CA: Academic Press.

Bjork, E. L., Bjork, R. A., & Anderson, M. C. (1998). Varieties of goal directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103 137). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123 144). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes* (pp. 35 67). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Camp, G., & deBruin, A. (2008). *What determines whether retrieval practice leads to forgetting or facilitation?* Paper presented at the 49th Annual Meeting of the Psychonomic Society, Chicago, IL.

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

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 facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Lan guage*, *61*, 153 170.

Chan, J. C. K. (2010). Long term effects of testing on the recall of nontested materials. *Memory*, *18*, 49 57.

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

Ciranni, M., & Shimamura, A. P. (1999). Retrieval induced forgetting in episodic memory. *Journal of*

*Experimental Psychology Learning, Memory and Cognition*, *25*, 1403 1414.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407 428.

Garcia Bajos, E., Migueles, M., & Anderson., M. C. (2009). Script knowledge modulates retrieval induced forgetting for eyewitness events. *Memory*, *17*, 92 103.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625 632). London: Academic Press.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259 284.

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*, 29 34.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval induced forgetting. *Psychological Science*, *12*, 148 152.

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

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*, 1 43.

Migueles, M., & Garcia Bajos, E. (2007). Selective retrieval and induced forgetting in eyewitness memory. *Applied Cognitive Psychology*, *21*, 1157 1172.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93 134.

Radvanksy, G. A. (1999). Memory retrieval and suppression: The inhibition of situation models. *Journal of Experimental Psychology: General*, *128*, 563 579.

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

Saunders, J., Fernandes, M., & Kosnes, L. (2009). Retrieval induced forgetting and mental imagery. *Memory & Cognition*, *37*, 819 828.

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval induced forgetting in an eyewitness memory paradigm. *Psychonomic Bulletin & Review*, *2*, 249 253.

Shivde, G., & Anderson, M. C. (2001). The role of inhibition in meaning selection: Insights from retrieval induced forgetting. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 175 190). Washington, DC: American Psychological Association.

Smith, E. E., Adams, N. E., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology*, *10*, 438 464.

Storm, B. C. (2010). 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*. Hove, UK: Psychology Press.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2005). Social metacognitive judgments: The role of retrieval induced forgetting in person memory and impressions. *Journal of Memory and Language 52*, 535 550.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2007). When intended remembering leads to unintended forgetting. *Quarterly Journal of Experimental Psychology*, *60*, 909 915.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval induced forgetting: The benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 230 236.

Storm, B. C., Bjork, E. L., Bjork, R. A., & Nestojko, J. F. (2006). Is retrieval success a necessary condition for retrieval induced forgetting? *Psychonomic Bulletin & Review*, *13*, 1023 1027.