Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 5, 1113–1124

0278-7393/11/$12.00 DOI: 10.1037/a0023549

Retrieval Experience as a Modifier of Future Encoding: Another Test Effect

Elizabeth Ligon Bjork
University of California, Los Angeles

Benjamin C. Storm
University of Illinois at Chicago

Research on how individuals monitor their level of comprehension during study paints a picture of learners as being insensitive to many of the factors or conditions of learning that can enhance long-term retention and transfer. In previous research, however, deWinstanley and Bjork (2004) demonstrated that learners—if made sensitive to the memorial benefits of generation in the context of an informative test following study of a text passage in which they had encoded both to-be-read and to-be-generated critical items—then became more effective processors of future to-be-read information presented in a 2nd text passage. In Experiments 1 and 2 of the present research, we explored the potential applicability of this effect by testing whether it could survive certain types of activity-filled delays. In Experiments 3 and 4, we tested whether enhanced processing of contextual information, an encoding strategy that could possibly have been discovered by participants during the testing episode for the 1st text passage, was a potential underlying cause of this effect. Together, our results bring to light an additional benefit of test taking and point to what might be considered necessary and sufficient conditions for leading learners to become more effective processors of future to-be-learned information.

Keywords: generation, metacognition, testing

The ability of individuals to monitor or judge their own learning during acquisition or study has been a subject of much recent research on metacognitive processes and, undoubtedly, much of the interest in this question stems from the assumption that such judgments play a critical role in determining how, as learners, we decide to allocate our future learning resources (e.g., R. A. Bjork, 1999; Jacoby, Bjork, & Kelley, 1994). On the basis of such judgments, for example, students might decide to review one chapter rather than another or to restudy one set of materials rather than study a new set of materials in preparation for an upcoming examination.

Several lines of evidence would seem to warrant the making of such an assumption. First, when individuals are asked to monitor their degree of learning by making judgments of learning (JOLs) during acquisition or initial study and, also, to make decisions about their future study behavior (e.g., which items they would like to restudy and for how long), they tend to choose for additional study those items that they judged to be more difficult (i.e., the items to which they gave low JOLs), resulting in a negative correlation between JOLs and study-choice and/or study-time allocation (e.g., Dunlosky & Hertzog, 1997; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; for a review, see Son & Metcalfe, 2000). Second, as recently demonstrated by Kornell and Metcalfe (2006) in a series of studies, learners perform better, under certain circumstances, when allowed to study items they have chosen rather than items they have not chosen. Finally, Metcalfe and Finn (2008)—by creating situations in which learners' metacognitions or JOLs were dissociated from their recall performance—have obtained compelling direct evidence that individuals use their JOLs in making decisions regarding their future study activities.

Although it seems safe to assume that individuals do use such judgments to guide their future study choices or activities, it is also clear from other research that individuals are often far from accurate in making these judgments, frequently suffering from illusions of comprehension (e.g., R. A. Bjork, 1999; Jacoby et al., 1994; Koriat, 1997, 1998). Learners, for example, can be led to think that their level of skill is greater than it actually is owing to conditions of learning (e.g., massed or blocked practice) that enhance or support performance during study or training but actually impair long-term retention and/or transfer (e.g., Simon & Bjork, 2001). Similarly, learners can be led to make JOLs that perfectly mismatch their later performance on a test by basing them on the fluency with which they can retrieve answers from long-term memory in the presence of cues that are available at the time of study but that will not be available at the time of a later test (Benjamin, Bjork, & Schwartz, 1998).

One account for the occurrence of such dissociations between JOLs and actual performance is offered by the new theory of disuse (R. A. Bjork & Bjork, 1992), which would posit that such dissociations occur when learners base their judgments on retrieval

This article was published Online First May 16, 2011.

Elizabeth Ligon Bjork, Department of Psychology, University of California, Los Angeles; Benjamin C. Storm, Department of Psychology, University of Illinois at Chicago.

The present research was supported in part by Collaborative Award WU-HT-09-15 from the James S. McDonnell Foundation. We thank Jeri L. Little and Nancy Sin for their help in conducting aspects of this research and members of Cogfog for helpful comments. Aspects of these results were presented at the 2006 annual meeting of the Western Psychological Association, Palm Springs, California, and at the 2008 annual meeting of the Psychonomic Society in Chicago, Illinois.

Correspondence concerning this article should be addressed to Elizabeth Ligon Bjork, Department of Psychology, University of California, Los Angeles, CA 90095-1563. E-mail: elbjork@psych.ucla.edu

strength (i.e., the current activation or accessibility of an item's representation in memory) rather than on storage strength (i.e., how entrenched or interassociated a memory representation is with related knowledge and skills). Because retrieval strength is a poor indicator of actual learning, it is also a poor indicator of long-term performance. Thus, conditions of learning that enhance retrieval strength as opposed to storage strength (e.g., blocked vs. distributed practice) can render learners inappropriately overconfident. From a slightly different perspective, Koriat (1997) posited that such dissociations occur because, during original study, learners are overly sensitive to intrinsic factors (e.g., the perceived association between cues and targets when both are present during study) while being relatively insensitive to extrinsic factors (e.g., the number of study repetitions and presentation durations, as well as encoding operations, e.g., level of processing or interactive imagery). This biased sensitivity is problematic because intrinsic factors often have less impact on future test performance than do extrinsic factors. In the present article, we report research concerned with the sensitivity of learners to the memorial benefits of one such extrinsic factor—the encoding operation of generation—and the question of whether, if made sensitive to these benefits, learners would then adopt more effective encoding strategies in the processing of new information. Stated more generally, we examined whether experiencing the memorial benefits of generative processing during a test can make learners more effective at learning in the future.

Generation as an Effective Condition of Learning

Learners tend to remember information that they take an active role in generating better than they remember information that is provided intact for them to study or read. For example, if learners generate the word *banana* from a word fragment (e.g., *b_n_n_*) as opposed to being given the intact word to read, they will recall it better on a later retention test. Or, if required to generate the exemplar *banana* to a category-plus-letter-stem cue (e.g., *fruit—ba___*) versus being given the intact pair to study, they will recall *banana* better in response to the cue *fruit* on a later test. This memorial benefit of generation (e.g., Jacoby, 1978; Slamecka & Graf, 1978) has proven both to be robust and to extend to a variety of learning materials, including lists of words, trivia questions (e.g., deWinstanley, 1995), and mathematical problems (e.g., McNamara & Healy, 1995a, 1995b; Pesta, Sanders, & Murphy, 1999). Moreover, recent work by Metcalfe and colleagues has indicated that the memorial benefits of a number of variables shown to enhance learning in the laboratory, including forced generation versus passive reading, transfer to learning in classroom settings. For example, in a 6-week program with sixth graders in a high-risk academic setting, their learning of vocabulary items critical for textbook understanding using a computer-based study program incorporating such principles as generation, multiple testing, and spaced practice was greatly enhanced compared with their learning from self-study sessions. (For a detailed review of this work, including studies conducted in both the classroom and the laboratory, see Metcalfe, 2006; Metcalfe & Kornell, 2007; Metcalfe, Kornell, & Son, 2007.)

Theoretical accounts of the generation effect have emphasized the critical nature of the relationship between encoding and retrieval processes. For example, the procedural account (Crutcher & Healy, 1989; McNamara & Healy, 1995a, 1995b) assumes that when learners are required to generate information at study, as opposed to reading it, they are more likely to use encoding procedures that can then be reinstated during a later retention test. When a later test does invoke such procedures, a generation advantage should occur; if not, a generation advantage should not occur.

Also emphasizing the critical nature of this relationship but from a slightly different perspective, the transfer-appropriate multifactor account (deWinstanley, Bjork, & Bjork, 1996)—built on the two-factor account of Hirshman and Bjork (1988) and the multifactor account of McDaniel, Waddill, and Einstein (1988)—assumes that the act of generation strengthens whatever type of information is used by the learner to complete the generation task and, thus, the consequence of the generation task for later memory performance depends on whether the information so enhanced is information to which a later test is sensitive. When there is a good match between these types of information, generation advantages should occur; when there is not, generation advantages should not occur.

To illustrate, if participants are being asked either to generate or simply to read targets in lists of cue–target pairs, the transfer-appropriate multifactor account would argue that the generation task could lead to enhanced strengthening of one or more of the following types of information depending on the specific nature of the generation task: target-specific information (information specific to the target item itself, such as how it looks), cue–target relational information (information about the specific relation that the target has to the cue—an antonym, synonym, rhymes with, etc.), and target–target or whole-list relational information (e.g., similarities among targets, such as a shared categorical membership). Accordingly, if a given generation task had specifically or primarily strengthened cue–target relational information, for example, then a later cued-recall test—assumed to be sensitive to such information (e.g., Einstein & Hunt, 1980; Tulving, 1962)—should reveal a generation advantage, whereas a later free-recall test—assumed to be sensitive to target–target or whole-list relational information (Einstein & Hunt, 1980; Tulving, 1962)—might not. Or the reverse pattern should occur when the generation task primarily strengthens target–target or whole-list relational information: Namely, a later free-recall test should reveal a generation advantage, whereas a later cued-recall test might not. In support of this account, changes in a variety of factors—such as the type of test learners expect, whether to-be-read or to-be-generated items are mixed together (i.e., between-subjects or within-subject manipulations of generation vs. read), and the specific requirements of the generation task combined with the type of memory test administered—have led to a continuum of outcomes ranging from large to small to no generation advantages (e.g., Begg, Vinski, Frankovich, & Holgate, 1991; deWinstanley & Bjork, 1997; deWinstanley et al., 1996; McDaniel, Riegler, & Waddill, 1990).

Making Learners Sensitive to Generation as an Effective Condition of Learning

In previous research, deWinstanley and Bjork (2004) explored the sensitivity of learners to the memorial benefits of generation and hypothesized that—if made sensitive to such benefits—they might adopt more effective encoding strategies in the processing of new information. The general research strategy they implemented in exploring this possibility, which we have also used in the present research, was as follows: Participants were first presented with a short passage to study of the type that would appear in an undergraduate introductory textbook but in which both to-be-generated and to-be-read critical items had been embedded. Next, participants' recall for these critical items was assessed by a fill-in-the-blank test (also known as a *cloze* test; Taylor, 1953). Then, after the experience of this test, a new text passage, also containing both to-be-generated and to-be-read critical items, was presented for study and followed by the same type of test for the critical items. Thus—before presentation of the second text passage for study—participants would have had the opportunity to engage in both the generating and the reading of critical items in a previous passage as well as the opportunity to experience a generation advantage in their own performance on the test of those items. Hence, if, as hypothesized, such an experience could be sufficient to induce participants to adopt a more effective way of encoding future to-be-read information, a generation advantage should be attenuated or possibly eliminated in the test of the second passage.

In two studies using this procedure, deWinstanley and Bjork (2004) obtained results consistent with this hypothesis. Specifically, although a generation advantage was observed in the test of the first passage, no generation advantage was observed in the test of the second passage. It is important to note, however, that the absence of a generation advantage on the second test did not occur at the expense of the generated items. Instead, recall of the to-be-read items presented in the second passage improved to the level of that for the to-be-generated items, which did not differ from their level of recall in the test of the first passage. In their second study, deWinstanley and Bjork also asked participants to describe what they had noticed about their performance on the first memory test. More than half of the participants produced responses that could be coded as noticing a generation advantage (e.g., "mostly remembered the words I had to figure out"), suggesting that participants were generally aware of the memorial advantage of generating compared with reading after the first test. It is important to note that participants who demonstrated such awareness were significantly less likely to exhibit a generation advantage in the test of the second paragraph than were participants who did not demonstrate such awareness, suggesting that participants were able to use their awareness of the generation advantage effectively to change their encoding strategy for the second paragraph.

Additionally, in two follow-up studies, deWinstanley and Bjork (2004) obtained results indicating the importance of the testing experience for leading participants to develop more effective encoding strategies. Specifically, when they denied participants the opportunity to experience the memorial advantage of generation versus reading in their own performance during a testing episode, by manipulating the requirement to read versus generate between passages (Experiment 3) or between participants (Experiment 4), recall of to-be-read critical items presented in a second passage remained significantly poorer than that for to-be-generated items.

Questions Addressed in the Present Research

The research of deWinstanley and Bjork (2004) indicates that learners, if given an informative test experience, can discover for themselves how to process or encode future to-be-learned information more effectively. Or, in the terms of Koriat (1997), they showed that making learners sensitive to the power of generation as a condition of learning can lead them, in turn, to adopt enhanced strategies for the encoding of new information via reading—that is, even for information that they are not required to generate. These findings thus raise many interesting questions: some regarding the underlying cause of the enhanced encoding strategies and some regarding the potential application of these findings to educational practices.

With respect to the potential applicability of these findings for educational purposes, certainly an important issue would be the effect's durability. In the studies of deWinstanley and Bjork (2004), the second passage was always presented with very little delay after the test of the first passage, raising the question of whether the testing experience only leads to enhanced encoding of new information when that information is presented for processing immediately after the test experience. Perhaps, for example, a delay filled with other attention-demanding activities inserted between the testing experience and the presentation of the next passage would have prevented participants from adopting a more effective processing strategy for subsequent to-be-read information. If the presumed effect of the testing experience cannot persist across a delay filled with other activities, then its potential applicability to the classroom would seem limited, and it was this question that we addressed in Experiment 1.

Experiment 1

In Experiment 1, although, just as deWinstanley and Bjork (2004) had previously done, we presented participants with two passages to study, each containing both to-be-generated and to-be-read critical items, we now inserted a delay filled with other activities between the testing of the first passage and the presentation of the second passage. If presentation of the second passage must follow more or less immediately after the testing episode in which participants experience the memorial benefit of generation for them to adopt improved encoding strategies for the to-be-read items in a second passage, then we would not expect to see the generation advantage diminished or eliminated in the test of the second passage. If, instead, this beneficial effect of the testing experience can survive a delay filled with other types of attention-demanding cognitive activities, then participants should go on to process the to-be-read items more effectively in the second passage—despite the delay in its presentation—and a generation advantage should be diminished or eliminated in the test of the second passage. That is, we should see the same general pattern of results as that observed by deWinstanley and Bjork in their first two experiments: a generation advantage on the test of the first passage with this advantage being diminished or eliminated on the test of the second passage.

Method

Participants. A total of 32 undergraduate students (26 women, six men) from the University of California, Los Angeles, participated for credit in an introductory psychology course.

Materials. The same passages used by deWinstanley and Bjork (2004, Experiment 1B) were used in the present Experiment 1: one on motivation and goal orientation and one on Bloom's

(1956) taxonomy of educational objectives, with the presentation order of the two passages counterbalanced across participants. Each passage was divided into 12 stand-alone sentences consisting of between 7 and 16 words, with the first two sentences serving as buffers; that is, the critical items contained in them were never tested. The remaining 10 sentences, however, each contained a single critical item that was later tested for retention. For half of these sentences, the critical item was kept intact, underlined, and colored red (i.e., the to-be-read items). For the other half of the sentences, the critical item was fragmented in addition to being underlined and colored red (i.e., the to-be-generated items). To illustrate, in the sentence, "The lowest level of cognitive learning objectives is the knowledge level," the word *knowledge* is the critical item and would have appeared in red print. When in the to-be-read condition, the word was presented intact (i.e., *knowledge*); when in the to-be-generated condition, it was presented with several of the vowels missing (i.e., *kn_wl_dg_*). The particular sentences containing to-be-read critical items and those containing to-be-generated critical items were counterbalanced across subjects. The generation task was designed to be successful and, in fact, participants were able to generate critical items successfully about 99% of the time.

Procedure. At the onset of the experiment, participants—all of whom were tested individually—were seated in front of a computer and given response booklets. They were then told that they would be presented with two passages to study, each of which would appear one sentence at a time on the computer screen and each of which would be followed by a memory test. As in the previous deWinstanley and Bjork (2004) experiments, the exact nature of the memory test was not described to them—only that they would receive a memory test for each passage. They were also instructed that each sentence would contain one item appearing in red print, that this item would be presented as either a word fragment or an intact word, and that they should write the solution to each red fragment or the intact red word on a separate page in their response booklet, turning the page after writing in preparation for the next phrase. Examples of generate and read items were provided, and participants were shown how to use the booklets properly.

After receiving the instructions, participants were told to press any key to begin presentation of the sentences, which appeared on the screen for 17 s each. Following presentation of the final sentence of the first passage, participants engaged in an unrelated 2-min distractor task, during which they tried to navigate a maze, and they were then given a memory test. During this test, sentences from the first passage were presented in the same order as they had been presented during study, except now all of the critical items were missing (e.g., "The lowest level of cognitive learning objectives is the ________ level"). Participants were given 2 min to recall as many of the missing critical items as they could.

Once they had completed the memory test, participants were given a booklet containing two unrelated distractor activities: one requiring them to estimate the number of blocks making up different three-dimensional figures and one requiring them to identify states of the United States from outlines of their shapes. After engaging in these tasks for 15 min, the second passage was presented and tested in the same manner as the first passage had been. Finally, after completing the second test, the participants were debriefed and thanked for their participation.

Results and Discussion

Correct recall percentages for critical items on the tests for the first and second passages are shown in the top panel of Table 1 and were analyzed using a 2 (read vs. generate) × 2 (first passage vs. second passage) repeated-measures analysis of variance (ANOVA). As expected, on the test of the first passage, the generated items were recalled significantly better than were the read items, $t(31) = 3.32$, $p < .001$, $d = 0.74$. Importantly, however, with respect to the critical issue being addressed by Experiment 1—whether the effect observed by deWinstanley and Bjork (2004) can survive a delay filled with other attention-demanding activities—a generation advantage was not found in the test of the second passage; instead, to-be-read items were recalled at roughly the same rate as to-be-generated items, $t(31) = 0.50$, $p = .619$, $d = 0.09$. Additionally, the interaction was statistically significant, $F(1, 31) = 7.23$, $MSE = 6.57$, $p = .011$, $\eta_p^2 = 0.19$, with planned-comparison *t* tests confirming that the recall of generated critical items did not differ significantly from the test of the first passage to that of the second passage, $t(31) = 0.71$, $p = .482$, $d = 0.13$, whereas the recall of the read critical items did improve significantly, $t(31) = 3.37$, $p = .002$, $d = 0.60$, indicating that the elimination of the generation effect in the second passage stemmed from an increase in performance for read items—not diminished performance for the generated items.

The results of Experiment 1 demonstrate that the beneficial effect of the testing experience—that is, its ability to lead learners to develop more effective encoding strategies for processing future information—can persist across at least a limited delay filled with other activities. A related issue in terms of the potential educational applications of this benefit, however, would be whether the test experience must occur more or less immediately after presentation of the passage in which participants both generated and read critical items. Or, might it be possible—as presumably would sometimes be necessary in educational settings—to delay the test without eliminating the learners' ability to benefit from the test experience?

We tested this possibility in Experiment 2 by using a procedure similar to that of Experiment 1. Again, we presented participants with two passages, each containing both to-be-generated and to-be-read critical items, but rather than inserting an activity-filled

Table 1
Correct Recall Percentages and Standard Errors for Critical To-Be-Generated and To-Be-Read Items in Tests of First and Second Passages in Experiments 1 and 2

	Paragraph tested			
	First		Second	
Critical item type	%	*SE*	%	*SE*
Experiment 1				
To be generated	58	4	62	4
To be read	38	5	59	4
Experiment 2				
To be generated	44	4	50	3
To be read	34	4	50	4

delay between the test of the first passage and the presentation of the second passage, we inserted such a delay between presentation of the first passage and its test. If participants need to experience the memorial benefit of generation in a testing episode that more or less immediately follows their study of the first passage to be led to adopt improved encoding strategies for a second passage, then we would not expect to see the generation advantage diminished or eliminated in the test of the second passage. If, instead, this beneficial effect of the testing experience does not require that the test occur immediately after presentation of the passage in which participants have first encoded critical items via generation or reading, then they should go on to process the second passage more effectively—despite a delay filled with other attention-demanding activities before the test of the first passage—and a generation advantage should be diminished or eliminated in the test of the second passage.

Experiment 2

Method

Participants. A total of 44 undergraduate students (18 women, 26 men) from the University of California, Los Angeles, participated in the experiment. Of these, 16 participants were paid for their participation and 28 received credit in an introductory psychology course.

Materials and procedure. The passages used—motivation and Bloom's (1956) taxonomy—were exactly the same as those used in Experiment 1, with the particular sentences containing to-be-read critical items and those containing to-be-generated critical items as well as the presentation order of the two passages counterbalanced across subjects. Additionally, all procedures used in conducting Experiment 2 were kept the same as those in Experiment 1. The only difference from Experiment 1 was the placement of the 15-min activity-filled delay. In Experiment 1, this delay occurred between the test of the first passage and the presentation of the second passage. In Experiment 2, it was placed between study of the first passage and its test. Then, following the delayed test of the first passage, the second passage was presented and followed by its test. The type of tests used was the same as that used in Experiment 1. Finally, as in Experiment 1, participants were not told the exact nature of the memory test at the start of the experiment—only that they would receive a memory test for each passage.

Results and Discussion

Correct recall percentages for critical items on the tests for the first and second passages are presented in the bottom panel of Table 1 and were analyzed using a 2 (read vs. generate) × 2 (first passage vs. second passage) repeated-measures ANOVA. As in Experiment 1, generated items were recalled significantly better than were read items in the test of the first passage, $t(43) = 2.46$, $p = .018$, $d = 0.37$. It is important to note, however, with respect to the question of interest, a generation advantage was not found in the test of the second passage, where generated items were recalled at exactly the same rate as read items, $t(43) = 0.00$, $p = 1.000$, $d = 0.00$.

Thus, the pattern of results obtained across Experiments 1 and 2 indicates that the beneficial effect of the testing experience originally observed in the studies of deWinstanley and Bjork (2004)—that is, its ability to lead learners to develop more effective encoding strategies for processing future information—can persist across at least a limited delay filled with other activities and, furthermore, does not require that the test be administered immediately after presentation of the first passage. Delays of considerably longer duration will, of course, need to be explored in future research. It may prove necessary, for example, to provide students with substantially more experience with generation strategies and their memorial consequences (than the few minutes provided in the present paradigm) for the effect to survive delays of educationally practical significance. Moreover, future research will need to determine both the boundary conditions for the persistence of this benefit and how best to structure educational materials and practices so as to take advantage of it. Nonetheless, the present results are at least consistent with the possibility of potential applications to the classroom setting.

Experiment 3

Whereas the present Experiments 1 and 2 explored the potential applicability of the testing benefit observed by deWinstanley and Bjork (2004) to instructional settings and obtained results consistent with such an application, they do not directly address the theoretical question of how participants are actually improving their encoding of information in the second passage. Although, as previously discussed by deWinstanley and Bjork (2004), there may be a number of ways to induce participants to improve their processing of to-be-read items, the specific question we addressed in Experiment 3 was how they were doing so in the present situation—that is, after being exposed to both types of processing and then experiencing the advantages of encoding by generation versus reading in the context of a testing event. One possibility underlying such improvement might be as follows: During original study, participants used contextual information provided by other words in the passage to help them complete or encode the to-be-generated critical items and, then, used this information again in the subsequent fill-in-the-blank test to aid their recall. Indeed, the use of such a strategy—that is, to use contextual information first to help complete and then to help recall the generated items—could underlie the generation advantages observed on the tests of the first passages in both the original deWinstanley and Bjork (2004) studies as well as the present experiments and, additionally, would be an explanation consistent with both the procedural (e.g., McNamara & Healy, 1995b) and the transfer-appropriate multifactor accounts of generation effects (e.g., deWinstanley et al., 1996; Hirshman & Bjork, 1988; McDaniel et al., 1988). Specifically, it would have been (a) the ability to reinstate during test the cognitive procedures used during study or (b) the match between the information strengthened while completing the generation task and the information needed to perform well on the later test that had resulted in the observed generation advantages.

Should this explanation for the generation advantage observed in these studies be correct, perhaps participants—becoming aware of both their superior recall of generated items and their use of such contextual information in recalling them on the test—then attended to such contextual information during the study of the second passage for both types of critical items, consequently eliminating a generation advantage in their performance on the test of the second passage. Such an account would also be consistent with deWinstanley and Bjork's (2004, Experiment 2) observation that the generation advantage was not eliminated on tests of the second passage when participants had only received to-be-generated critical items during study of the first passage. Even if these participants were using contextual information in the same way as we are proposing while they studied the first passage, it may have been difficult for them to notice the role of this strategy in aiding their recall during the test because they were only recalling items they had generated and, thus, were not able to experience a contrast between their ability to recall words encoded via generation versus their ability to recall words encoded via reading. Consequently, they would have been less likely to transfer the use of this strategy when encoding to-be-read critical items presented in the second passage. Such a failure would also be consistent with previous research indicating that learners are not able to judge the efficacy of an encoding strategy during its execution and are unlikely to switch from a less to a more effective strategy without an opportunity to experience the strategies' relative effectiveness (e.g., Brigham & Pressley, 1988; Dunlosky & Hertzog, 2000; Shaughnessy, 1981).

We explored this potential explanation for improved encoding of to-be-read critical items in the second passage by giving participants a different type of test following presentation of the first passage in Experiment 3, namely, one that would not provide them with specific contextual information during the testing process—that is, a test that would not re-present the exact words or context that had previously surrounded the to-be-generated or to-be-read critical items during the study phase. Our reasoning in so doing was as follows: If this explanation is correct, then when the test following study of the first passage does not provide such specific contextual information, the testing experience should not lead participants to the discovery of this encoding strategy and, consequently, the generation advantage should not be eliminated in the testing of subsequently presented material. Thus, in Experiment 3, we followed presentation of the first passage with a test that did not provide such information—specifically, a free-recall test. If the proposed explanation for improved processing of information in the second passages of the deWinstanley and Bjork (2004) studies as well as those of the present Experiments 1 and 2 is correct, then a generation advantage should be observed in both the test of the first passage and that of the second.

Experiment 3

Method

Participants. A total of 44 undergraduate students (32 women and 12 men) from the University of California, Los Angeles, participated in the experiment for credit in an introductory psychology course.

Materials and procedure. Although the same materials and counterbalancing procedures as were used in Experiments 1 and 2 were used in Experiment 3—that is, the motivation and Bloom's (1956) taxonomy passages, with the particular sentences containing to-be-read critical items and those containing to-be-generated critical items as well as presentation order of the two passages counterbalanced across participants—two important differences in the procedure were introduced. First, no delays of 15 min were used at any point in Experiment 3. Rather, we used a filled delay of 2 min between the presentation and test of both passages, during which participants tried to navigate a maze, and only an unfilled delay of approximately 30 s between the test of the first passage and the presentation of the second passage, which replicated the schedule of delays used by deWinstanley and Bjork (2004). Second and most important, the nature of the test of the first passage was altered. Rather than completing a fill-in the-blank test following study of the first passage—as was done in all the previous studies of deWinstanley and Bjork as well as the present Experiments 1 and 2—participants completed a free-recall test for which they were instructed to write down as many of the critical terms that had appeared in red print as possible, including both intact words and words that they had to complete. In contrast, the format of the test for the second passage was identical to the fill-in-the-blank type of test as used in Experiments 1 and 2. As in both Experiments 1 and 2, the exact nature of the memory tests that participants would receive was not described to them at the start of the experiment—only that they would receive a memory test for each passage.

Results and Discussion

Importantly for our ability to examine whether a generation advantage would persist in the test of the second passage, correct recall performance on the free-recall test of the first passage exhibited a substantial effect of generation, with generated items ($M = 53\%$, $SE = 3\%$) recalled significantly better than read items ($M = 22\%$, $SE = 2\%$), $t(43) = 6.56$, $p < .001$, $d = 1.00$. Furthermore, with respect to the primary question addressed by Experiment 3, a generation effect was also found for the fill-in-the-blank test of the second passage, with the generated items ($M = 38\%$, $SE = 4\%$) being recalled significantly better than the read items ($M = 29\%$, $SE = 4\%$), $t(48) = 2.25$, $p = .020$, $d = 0.34$.

These results are thus consistent with our suggestion of the possible role played by the processing of contextual information—both during initial generation and then later during the testing experience—for leading participants to improve their future encoding strategies. Moreover, they suggest that merely experiencing the benefits of generation is not sufficient to lead participants to adopt better encoding strategies for future to-be-learned information, as indicated by the generation advantage persisting in their recall performance on the test of the second passage, an issue to which we return in the General Discussion.

Although we hoped to find a generation advantage in the free-recall test of the first passage in Experiment 3, as otherwise we would not have been able to test if the generation advantage was eliminated or diminished in the test of the second passage, this necessity could have been a potential flaw in our research strategy given that previous studies using heterogeneous lists of cue–target pairs have typically found only marginal or minimal generation advantages for free-recall as opposed to cued-recall tests (e.g., Hirshman & Bjork, 1988; Slamecka & Katsaiti, 1987). When, however, targets are selected from taxonomic categories and the categorical nature of the list is made apparent to learners by using the name of the category as the cues, significant generation effects have been obtained in free recall (e.g., Begg, Snider, Foley, & Goddard, 1989; McDaniel et al., 1988; also see deWinstanley & Bjork, 1997). Although constructs developed in theoretical accounts of generation effects found with lists of individual items and/or lists of cue–target pairs (as discussed in the introduction) may not be completely applicable to explaining generation effects found with coherent text passages, one possibility is that a coherent text passage functions similarly to a categorized list of cue–target pairs, leading participants to engage in something akin to whole-list relational processing to which free-recall tests are assumed to be sensitive (e.g., Einstein & Hunt, 1980; Tulving, 1962). Perhaps, as with categorized lists of cue–target pairs, this type of processing, in addition to the enhanced item-specific processing that the to-be-generated items would have incurred, contributed to our finding of a generation advantage in the free-recall test.

Although the results of Experiment 3 are definitely consistent with our proposed explanation of how participants are led to adopt improved future processing strategies, they do not directly speak to the enhanced processing of contextual information during encoding of the second passage as an important factor underlying this effect. Thus, to test our hypothesis more directly, we conducted a fourth experiment in which we varied both the nature of the test given to participants following their study of the first passage—that is, tests that did or did not provide specific contextual information during the testing episode—and the type of information for which we tested following the second passage—specifically, testing for contextual information rather than critical items. Our reasoning in so doing was that it would only be participants who could discover this strategy during the testing experience (i.e., those given a test for the first passage that included contextual information) who would then go on to process such information more effectively for to-be-read critical items as well as to-be-generated items in the second passage. Consequently, these participants should reveal a superior ability to recall contextual items in the test of the second passage compared with those participants not afforded this opportunity.

Experiment 4

Method

Participants. A total of 48 undergraduate students from the University of California, Los Angeles, participated in the experiment for credit in an introductory psychology course.

Materials and procedure. As in Experiment 3, participants once again studied two passages—motivation and Bloom's (1956) taxonomy—each containing to-be-generated and to-be-read critical items, with the particular sentences containing to-be-read critical items and those containing to-be-generated critical items as well as presentation order of the two passages counterbalanced across participants. Also, as in Experiment 3, a filled delay of only 2 min occurred between presentation of each passage and its test and only an unfilled delay of approximately 30 s occurred between the test of the first passage and presentation of the second passage, again replicating the original schedule of delays used by deWinstanley and Bjork (2004). And, as in Experiments 1, 2, and 3, the exact nature of the memory tests that participants would receive was not described to them at the start of the experiment—only that they would receive a memory test for each passage.

In Experiment 4, however, the format of the first test was randomly manipulated between participants: Half were given a fill-in-the-blank test and the other half were given a free-recall test. It is important to note that a new type of test for the second passage was given to all participants. Rather than being asked to recall the to-be-read and to-be-generated critical items from the second passage, participants were asked to recall a surrounding context word from each of the sentences. More specifically, each previously studied sentence was presented intact to the participants except for a single word left blank. The missing word was not the critical item that had been copied or generated during study; instead, it was a context word that presumably would have helped participants generate the incomplete critical item during study. To illustrate, for participants given the Bloom (1956) passage as their second paragraph to study, one of the sentences presented was "The lowest level of the cognitive learning objectives is the knowledge level," with the word *knowledge* as the critical item either to be generated or to be read and appearing in red print. Then, in the following fill-in-the-blank test, the corresponding sentence would be "The lowest level of the c_____ learning objectives is the knowledge level," with the word *cognitive* as the context word to be recalled.

Finally, at the end of the experiment, participants were asked questions to try to assess whether they had noticed anything different between the way they studied the first and the second passages and whether, after the first passage, they had developed a strategy for learning the information in the second passage better.

Results and Discussion

Recall performance for first-passage critical items. Correct recall percentages for to-be-generated and to-be-read critical items are presented in the top panel of Table 2 as a function of the type of test administered following study of the first passage. Correct recall performance was analyzed using a 2 (read vs. generate) × 2 (free-recall vs. fill-in-the-blank test) mixed-design ANOVA, with type of test serving as the only between-subjects variable. A main effect of generation was observed in the recall of critical words, $F(1, 46) = 10.47$, $MSE = 13.50$, $p = .002$, $\eta_p^2 = .19$, and, furthermore, the generation advantage did not vary depending on the type of test administered, $F(1, 46) = 0.13$, $MSE = 0.17$, $p = .721$, $\eta_p^2 = .00$, with both groups showing a significant generation

Table 2
Correct Recall Percentages and Standard Errors in Experiment 4 for Critical Items in First Passage and Context Words in Second Passage, Respectively, as a Function of Test 1 Type

	Test 1 type			
	Fill in the blank		Free recall	
Critical item type	%	*SE*	%	*SE*
First passage (critical items)				
To be generated	35	5	49	5
To be read	22	4	33	4
Second passage (context words)				
To be generated	45	5	33	5
To be read	43	5	30	5

advantage. That is, both those participants taking a free recall test and those taking a fill-in-the-blank test recalled significantly more critical items in the generation condition than in the read condition. Finally, recall performance was higher for participants taking a free-recall test than for participants taking a fill-in-the-blank test, $F(1, 46) = 6.68$, $MSE = 9.38$, $p = .013$, $\eta_p^2 = .13$.

Although our finding of a generation advantage for first-passage critical items following either type of test was of most importance for our being able to test the primary hypothesis of Experiment 4—namely, that context words would be better remembered by participants given a fill-in-the-blank test of the first passage than by participants given a free-recall test of the first passage—our finding of better overall recall performance for participants taking a free-recall test than for participants taking a fill-in-the-blank test is also of interest, given that better performance is typically observed on cued-recall versus free-recall tests. We believe, however, that we most likely observed this effect in the present experiment because the criterion for correct performance could be considered more stringent for our cued-recall (i.e., fill-in-the-blank) test than for our free-recall test. Specifically, participants receiving a free-recall test were given credit for any critical item they recalled, whereas participants receiving a fill-in-the-blank test had to place a recalled critical item in a particular sentence to receive credit. Indeed, if participants were given credit for any critical item recalled regardless of where that item was placed, the overall correct performance of participants receiving a fill-in-the-blank test increases to 39%, which is not significantly different from the overall correct performance of those participants receiving a free-recall test (41%). It is interesting that of the additional 24 critical items counted as correct with this relaxation of the scoring criterion on the fill-in-the-blank test, twice as many (16) were critical items that had been read versus ones that had been generated (eight). That is, twice as many to-be-read critical items were placed in an incorrect sentence as were to-be-generated items—consistent with the notion that the encoding of critical items via generation leads to a stronger link being formed between the critical item and the surrounding context words than does the encoding of critical items via reading.

Recall performance for second-passage context words. Having observed significant generation advantages for either type of test of the first passage, we could now test our primary hypothesis by comparing recall performance of context words in the second passage as a function of whether participants had received a free-recall or a fill-in-the-blank test of the first passage, which we analyzed using a 2 (read vs. generate) × 2 (free-recall vs. fill-in-the-blank test) mixed-design ANOVA. The free-recall versus fill-in-the-blank manipulation refers to the type of test participants experienced following study of the first passage, and the read versus generate manipulation refers to whether the to-be-tested context words had been presented in a sentence with a read or generated critical item, and the corresponding correct recall percentages are shown in the bottom panel of Table 2.

First, as indicated in Table 2, there was no overall effect of the read versus generate manipulation, $F(1, 46) = 0.60$, $MSE = 0.38$, $p = .441$, $\eta_p^2 = .01$, and, furthermore, planned comparisons revealed that the generation effect for context words was not significant in either the free-recall, $t(23) = 0.33$, $p = .747$, $d = 0.07$, or the fill-in-the-blank condition, $t(23) = 0.85$, $p = .405$, $d = 0.17$.

Of primary importance to our hypothesis, however, participants who had experienced a fill-in-the-blank test for the first passage ($M = 44\%$, $SE = 4\%$) significantly outperformed participants who had experienced a free-recall test for the first passage ($M = 32\%$, $SE = 4\%$), $F(1, 46) = 4.76$, $MSE = 9.38$, $p = .034$, $\eta_p^2 = .09$. Furthermore, this advantage did not vary depending on whether the critical item had appeared in a sentence containing a to-be-generated or a to-be-read critical item during processing of the second passage, $F(1, 46) = 0.07$, $MSE = 0.04$, $p = .797$, $\eta_p^2 = 0.00$. As expected and shown in the first row of the bottom panel of Table 2, context words associated with to-be-read critical items—that is, ones appearing in sentences with to-be-read critical items—were better recalled if participants had experienced a fill-in-the-blank test for the first passage than if they had experienced a free-recall test for the first passage. Likewise, context words associated with to-be-generated critical items—that is, ones appearing in sentences with to-be-generated critical items—were also better recalled if participants had experienced a fill-in-the-blank test for the first passage than if they had experienced a free-recall test for the first passage. These results thus strongly support our hypothesis that at least one factor underlying the elimination of a generation advantage in tests of the second passage is the enhanced processing of contextual information during participants' encoding of the second passage—a strategy arising from metacognitive processes occurring during a testing experience indicating the potential usefulness of attending to contextual information during initial encoding.

For participants receiving a fill-in-the-blank test for the first passage, the lack of a generation advantage for context words in the test of the second passage is consistent with our proposed hypothesis concerning at least one way in which these participants might be improving their processing of to-be-read critical items in the second paragraph. For participants receiving a free-recall test for the first passage, however, one might have expected to see a generation advantage for context words in the test of the second passage, given that a generated advantage had persisted in the recall test of the second passage following a free-recall test of the first passage in Experiment 3. Yet, in some previous research using lists of cue–target pairs as study materials, the finding of generation advantages for cue or context words has been somewhat evasive, with some studies finding no advantage (Begg et al., 1989), only a slight advantage (Slamecka & Graf, 1978), or only marginally significant advantages when summed across experiments (Greenwald & Johnson, 1989). In contrast, also using lists of cue–target pairs, McDaniel and Waddill (1990) found significant generation advantages for the context or cue words in two separate studies.

As suspected by McDaniel and Waddill (1990), a potentially important reason for the different outcomes of these two collections of studies was the use of intentional versus incidental learning conditions. In those studies failing to find generation effects for context words, participants were told to expect a memory test and were thus learning under intentional conditions, whereas McDaniel and Waddill took great pains to ensure their participants were learning under incidental conditions, eliminating from their analyses the data of any participants revealing that they had anticipated a memory test during study. In that all participants in the present research studied the passages under conditions of intentional learning, the present finding of no generation advantages for context words is thus consistent with the previous findings in the literature.

Additionally, in the present studies, it seems possible that once participants experienced a free-recall test for just the critical items presented in the first passage, they then paid less attention to surrounding context words in the second passage and, even if using them to help generate critical items when needed, they then focused all of their rehearsal activities on just the critical words. Such a strategy on the part of the free-recall participants would also seem consistent with the findings that not only did the fill-in-the-blank participants recall significantly more context words than did the free-recall participants but also that this overall difference was rather large (43% vs. 30%, respectively). These conjectures also seem consistent with the analysis of the metacognitive judgments of the participants, which we present in the next section.

Metacognitive judgments of participants. Following their completion of the experiment, participants were asked questions to assess whether they (a) had noticed anything different between the way they studied the first and second passages and (b) had developed a strategy for learning the information in the second passage better. As essentially all participants combined their answers to these two questions, we coded their answers together and assigned them to categories according to whether the participant gave any indication of (a) having switched to focusing more on context as well as the red items in study of the second passage, (b) having switched to focusing predominantly or only on the red words, or (c) having switched to a mixture of these two strategies. Responses that indicated some other strategy or were unclear as to whether the participant had engaged in any switching of strategy were assigned to the category of (d) unclear or other. Examples of responses indicating having switched to focusing more on context in the second passage are "For first passage, paid more attention to the red words; for second, looked more at context of sentence"; "read whole sentence more"; "looked for connections with surrounding words"; and "tried to pay more attention to how red word related to rest of sentence." Examples of responses indicating a switch to focusing more on the red words are "For second passage, tried to remember red words and didn't read paragraph"; "paid more attention to red words . . . rather than sentence as a whole"; and "paid more attention to red words and kept repeating red words in my mind." Examples of responses classified as a mixture are "In second passage, focused on red words and also on concepts" and "focused on overall sentence and repeated red words in my head." Examples of responses categorized as unclear or other are "practice made it easier" and "second time, things stood out more."

Of the 24 participants receiving a fill-in-the-blank test after the first passage, 14 gave responses that could be classified as indicating a switch in strategy to focusing more on context in the second passage, three indicated an increased focus on red words only, five were assigned to the mixture category, and six were assigned to the other category. Thus, the majority of these participants (58%) gave responses indicating a realization of the importance of developing a strategy for encoding the context in which the critical item was embedded—not just the critical items themselves. Furthermore, consistent with our hypothesis, these participants did correctly recall more context words ($M = 48\%$) on the second test than did those participants indicating a switch to focusing on red words ($M = 27\%$). It is interesting that the five participants who indicated that they adopted a mixture of the two strategies (i.e., encoding context but also rehearsing red words) also recalled context words less well (30%) than did the participants only indicating a change to enhanced context encoding.

Of the 24 participants receiving a free-recall test after the first passage, only three gave responses indicating a switch to focusing more on context, whereas 18 gave responses indicating a switch to focusing more or only on red words; three were assigned to the mixture category, and no responses fell into the other category. Although, as with the above comparisons, the number of observations are very small and thus need to be interpreted cautiously, it is interesting that the six participants whose responses fell into either the focus-on-context or the mixture category correctly recalled more context words (43% and 53%, respectively) than did those whose responses fell into the focus-on-red category (25%).

On the basis of the analysis of these responses, it seems clear that participants' metacognitions arising from the testing experience led the majority of them to switch to different strategies for encoding the information presented in the second passage. Participants experiencing the memorial benefit of generation during the fill-in-the-blank test predominantly reported developing a strategy for enhanced encoding of context words as well as critical items in the second passage, whereas the majority of participants experiencing the memorial benefit of generation in a free-recall test did not seem to become aware of any benefit that might have accrued from enhanced processing of context words as part of generating critical items.

General Discussion

Previously, deWinstanley and Bjork (2004) demonstrated that learners, if given an informative test experience, could discover for themselves how to become more effective processors or encoders of future to-be-learned information (see also E. L. Bjork, deWinstanley, & Storm, 2007). In particular, they showed that learners could be induced to develop more effective encoding strategies after experiencing the advantages of encoding by generation versus reading in the context of a testing episode—or, in the terms of Koriat (1997), after being made sensitive to the power of generation as a condition of learning. In the present research, we explored two important issues regarding these findings: the potential applicability of this effect for the enhancement of educational practices and a potential underlying cause of the improved processing.

With respect to the first issue, we addressed whether the effect could survive a delay filled with other attention-demanding activities. Specifically, in Experiment 1, we tested whether it could sustain an activity-filled delay between the testing experience of the first passage and the presentation of the second passage and, in Experiment 2, whether it could sustain an activity-filled delay between the presentation of the first passage and the testing of that same passage. Although much longer delays will need to be explored in future research, the results obtained in the present experiments are at least consistent with a potential application of the present findings for educational purposes in demonstrating that the effect of the testing experience—that is, its ability to lead learners to develop more effective encoding strategies for processing future information—can persist across a delay filled with other attention-demanding activities and, furthermore, does not require that the test be administered immediately after presentation of the passage in which learners first experience the process of encoding information via generation versus reading.

The second issue we addressed in the present research concerned the process by which participants—having been made sensitive to the mnemonic benefit of generation as a condition of learning (Koriat, 1997) via a testing experience—were then able to improve their encoding of information in a second passage. Our hypothesis was that they discovered, during the testing episode, the importance of paying attention to contextual information when generating critical items and then used this strategy during the second passage when encoding critical items, whether those items were presented as to be generated or to be read. In Experiment 3, we tested this hypothesis by giving participants a free-recall test instead of a fill-in-the-blank test following the presentation of the passage in which they had first encoded critical items via generation or reading. Because the free-recall test did not provide the same degree of contextual information during the testing process as did the fill-in-the-blank test (i.e., the exact words previously surrounding critical items), we assumed that such a testing experience would be less likely to lead participants into discovering this potentially beneficial encoding strategy. And, indeed, participants continued to show a generation advantage in their recall performance for critical items presented in the second passage.

Finally, in Experiment 4, we more directly tested whether adoption of such a strategy might have led to participants' enhanced encoding of the second passage. Specifically, we manipulated the degree of contextual information provided by the test of the first passage (i.e., free-recall vs. fill-in-the-blank tests) and then tested for context items rather than critical items in the test of the second passage. The idea was that only the participants who were likely to discover this strategy during the testing experience (i.e., those given the fill-in-the-blank test) would then go on to process such information more effectively in the second passage, and this expectation was borne out in the obtained results. Specifically, participants given a fill-in-the-blank test for the first passage recalled significantly more context items from the second passage than did participants given a free-recall test for the first passage.

Additionally, the results obtained across Experiments 3 and 4 point to what might be thought of as necessary and sufficient conditions for learners, given the experience of generating versus reading, to develop more effective strategies for the encoding of future to-be-learned information. First, it would seem that learners need to engage in generation and reading in the same encoding or study episode and, then, to experience the memorial benefits of generation over reading in a testing episode. The apparent importance of both of these experiences for producing strategy changes for future learning is thus consistent with other work in the literature indicating that individuals do not switch from a less effective strategy to a more effective one without an opportunity to experience their relative effectiveness (e.g., Brigham & Pressley, 1988; Shaughnessy, 1981, and, in particular, the framework of knowledge updating proposed by Dunlosky & Hertzog, 2000). As indicated by the results of Experiment 3, however, these two experiences would not appear to be sufficient, as these participants—although having both experiences—did not improve their processing of future to-be-read information. Thus, it would seem that the testing episode must also afford the opportunity for learners to realize a way in which the generation task might be enhancing their later memory performance—in the present situation, by tying the to-be-learned critical item to the surrounding contextual information in the sentence, that is, the context that is helping them both to generate the critical item and to define its meaning. In contrast, when given a free-recall test—although such a test may allow learners to experience the memorial advantage of generation—such a testing experience does not seem to inform them as to how to take advantage of such an insight to improve their encoding of information in the second passage. In general, then, it would seem that the testing experience has to be informative regarding how to use the metacognitive insights gained during the testing episode for such an experience to lead to the development of improved encoding strategies for future to-be-learned information.

Concluding Comments

As has clearly been documented in previous research, it is not just during study that learning takes place: Learning can also occur during tests. Indeed, much laboratory research (e.g., Carrier & Pashler, 1992; Landauer & Bjork, 1978; Roediger & Karpicke, 2006) has demonstrated the power of tests as learning events and, moreover, has shown that a test, even when no corrective feedback is given, can be considerably more effective for the long-term retention of material than additional study of it. Such observed benefits of testing are presumed to occur because the act of retrieving information modifies its representation in memory so as to make it more recallable in the future (R. A. Bjork, 1975). In addition to such specific effects of tests on learning, however, we believe that the present research demonstrates that another type of learning can also take place during tests—in particular, a higher order type of learning, such as the learning of an improved strategy for encoding future information. Furthermore, although we have focused on only one such strategy in the present research—that engendered by the generation and testing of to-be-learned information—it seems possible that learners could be made sensitive to other types of encoding operations or conditions of learning that enhance long-term performance through similar testing experiences. Thus, it seems to us that the present line of research paints a promising picture from an applied perspective: Namely, providing students with opportunities to experience the consequences of differentially effective encoding processes in their own performance—either in the context of tests, as was done in the present research, or perhaps in other ways as well—can lead students to discover and then to adopt on their own more effective ways of processing future to-be-learned information. That is, beyond enhancing the learning of the specific information being tested, testing episodes may provide a more general opportunity to learn how to learn more effectively in the future.

References

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 977–989. doi:10.1037/0278-7393.15.5.977

Begg, I., Vinski, L., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & Cognition, 19,* 487–497. doi:10.3758/BF03199571

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68. doi:10.1037/0096-3445.127.1.55

Bjork, E. L., deWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can experiencing the outcome of different encoding strategies enhance subsequent encoding? *Psychonomic Bulletin & Review, 14,* 207–211. doi:10.3758/BF03194053

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). 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. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals* (Vol. 1). New York, NY: Longman.

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in older and younger adults. *Psychology and Aging, 3,* 249–257. doi:10.1037/0882-7974.3.3.249

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642. doi:10.3758/BF03202713

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 669–675. doi:10.1037/0278-7393.15.4.669

deWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychonomic Bulletin & Review, 2,* 538–541. doi:10.3758/BF03210990

deWinstanley, P. A., & Bjork, E. L. (1997). Processing instructions and the generation effect: A test of the multifactor transfer-appropriate processing theory. *Memory, 5,* 401–422. doi:10.1080/741941392

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32,* 945–955. doi:10.3758/BF03196872

deWinstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer appropriate processing. *Memory, 4,* 31–48. doi:10.1080/741940667

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journals of Gerontology, Series B: Psychological and Social Sciences, 52,* P178–P186. doi:10.1093/geronb/52B.4.P178

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462–474. doi:10.1037/0882-7974.15.3.462

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 588–598. doi:10.1037/0278-7393.6.5.588

Greenwald, A. G., & Johnson, N. M. S. (1989). The generation effect extended: Memory enhancement for generation cues. *Memory & Cognition, 17,* 673–681. doi:10.3758/BF03202628

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484–494. doi:10.1037/0278-7393.14.3.484

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667. doi:10.1016/S0022-5371(78)90393-6

Jacoby, L. L., Bjork, R. A., & Kelley, C. M. (1994). Illusions of comprehension, competence, and remembering. In D. Druckman & R. A. Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp. 57–81). Washington, DC: National Academy Press.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A. (1998). Illusions of knowing: A window to the link between knowledge and metaknowledge. In V. Y. Yzerbyt, G. Lories, & B. Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 16–34). London, England: Sage.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 609–622. doi:10.1037/0278-7393.32.3.609

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

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18,* 196–204. doi:10.3758/BF03197095

McDaniel, M. A., Riegler, G., & Waddill, P. J. (1990). Generation effects in free recall: Further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 789–798. doi:10.1037/0278-7393.16.5.789

McDaniel, M. A., & Waddill, P. J. (1990). Generation effects for context words: Implications for item-specific and multifactor theories. *Journal of Memory and Language, 29,* 201–211. doi:10.1016/0749-596X(90)90072-8

McDaniel, M. A., Waddill, P. J., & Einstein, P. J. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521–536. doi:10.1016/0749-596X(88)90023-X

McNamara, D. S., & Healy, A. F. (1995a). A generation advantage for multiplication skill training and nonword vocabulary acquisition. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 132–169). Thousand Oaks, CA: Sage.

McNamara, D. S., & Healy, A. F. (1995b). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory and Language, 34,* 399–416. doi:10.1006/jmla.1995.1018

Metcalfe, J. (2006). Principles of cognitive science in education. *APS Observer, 19,* 27.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15,* 174–179. doi:10.3758/PBR.15.1.174

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review, 14,* 225–229. doi:10.3758/BF03194056

Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology, 19,* 743–768. doi:10.1080/09541440701326063

Nelson, T. O., Dunlosky, J., Graf, E. A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207–213. doi:10.1111/j.1467-9280.1994.tb00502.x

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain" effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686. doi:10.1037/0278-7393.14.4.676

Pesta, B. J., Sanders, R. E., & Murphy, M. D. (1999). A beautiful day in the neighborhood: What factors determine the generation effect for simple multiplication problems? *Memory & Cognition, 27,* 106–115. doi:10.3758/BF03201217

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Shaughnessy, J. J. (1981). Memory monitoring accuracy and modification of rehearsal strategies. *Journal of Verbal Learning and Verbal Behavior, 20,* 216–230. doi:10.1016/S0022-5371(81)90389-3

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907–912. doi:10.1037/0278-7393.27.4.907

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. doi:10.1037/0278-7393.4.6.592

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language, 26,* 589–607. doi:10.1016/0749-596X(87)90104-5

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 204–221. doi:10.1037/0278-7393.26.1.204

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly, 30,* 415–433.

Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review, 69,* 344–354. doi:10.1037/h0043150

Received May 22, 2010
Revision received February 22, 2011
Accepted March 1, 2011 ■