

Improving encoding strategies as a function of test knowledge and experience

Benjamin C. Storm¹ · Michelle L. Hickman¹ · Elizabeth L. Bjork²

Published online: 28 January 2016
© Psychonomic Society, Inc. 2016

Abstract Information that is produced or generated during learning is better remembered than information that is passively read, a phenomenon known as the generation effect. Prior research by deWinstanley and Bjork (*Memory & Cognition*, 32, 945–955, 2004) has shown that learners, after experiencing the memorial benefits of generation in the context of a fill-in-the-blank test following the study of a text passage containing both to-be-read and to-be-generated items, become more effective encoders of to-be-read items on a second passage, thus eliminating the generation effect on a subsequent memory test. Current explanations of this phenomenon assume that learners need to actually experience the generation advantage on the test of the first passage to become more effective encoders of to-be-read items on the second passage. The results of the present research, however, suggest otherwise. Although experiencing a test of the first passage does appear to be critical for leading participants to become better encoders on the second passage, experiencing a generation advantage on the test for the first passage is not. More generally, these results shine new light on the generation effect as well as how and why taking tests has the potential to improve subsequent learning.

Keywords Generation · Metacognition · Reading · Testing

✉ Benjamin C. Storm
bcstorm@ucsc.edu

¹ Department of Psychology, University of California, Social Sciences II, Room 277, 1156 High Street, Santa Cruz, CA 95064, USA

² University of California, Los Angeles, CA, USA

Learning activities that require an active contribution from the learner can produce a greater memorial advantage compared to activities that are more passive, such as reading or studying provided information. If learners, for example, are required to generate information during study (say, generating words from subsets of their letters), they tend to perform better on a later recall test for such information than do learners who simply read that same information presented intact during study. This memorial advantage for generation over reading—referred to as the generation effect (Jacoby, 1978; Slamecka & Graf, 1978; for a review, see Bertsch, Pesta, Wiscott, & McDaniel, 2007)—has been demonstrated to be a robust phenomenon, occurring for a variety of materials and in educational contexts (e.g., deWinstanley, 1995; McNamara & Healy, 1995a, b; Metcalfe & Kornell, 2007; Metcalfe, Kornell, & Son, 2007; Pesta, Sanders, & Murphy, 1999).

Nonetheless, there are also times when requiring learners to encode information via generation (versus reading) does not result in a memorial advantage for the generated information. Thus, an adequate theoretical account of the generation effect must be able to predict both the conditions under which it should and should not be observed. A key feature of theoretical accounts that are able to make such predictions is the assumption that the relationship between encoding and retrieval processes is a critical determiner of its occurrence—somewhat similar to the assumption of transfer-appropriate processing (Morris, Bransford, & Franks, 1977).

The procedural account of the generation effect, for example, states that the effect is observed when the learner is able to reinstate the particular procedures that took place during generation on a later retention test (Crutcher & Healy, 1989; McNamara & Healy, 1995a, b). The transfer-appropriate multifactor account (deWinstanley, Bjork, & Bjork, 1996), which is built upon the two-factor (Hirshman & Bjork, 1988) and multifactor (McDaniel, Waddill, & Einstein, 1988) accounts,

asserts that generation effects occur when the act of generation strengthens information that will be useful for correct performance on a subsequent test. That is, based on these two accounts, when the later retention or criterion test is sensitive to the generated information, or evokes the same procedures as those practiced or learned during the generation task, then a generation advantage should be observed; when it does not, however, a generation advantage should not be observed.

Experiencing the generation effect enhances subsequent encoding

Implementing generation during learning may provide benefits beyond that of improving recall of the material being generated. Using a multitrial design, for example, deWinstanley and Bjork (2004) found that when participants experience the memorial benefits of generation for learning on one trial, that experience seemed to lead to enhanced encoding of new information on a future trial. That is, they found that experiencing a memorial advantage for the generated information presented in an initial text passage led participants to modify their encoding strategies such that they then learned a second passage better than if they had not experienced a generation advantage.

In the paradigm employed by deWinstanley and Bjork (2004), participants were presented with a first text passage to learn, with the passage presented one sentence at a time. Each sentence contained a critical target word (printed with red letters) that was either presented intact (to-be-read targets) or with some letters missing (to-be-generated targets). For the generated targets, participants used what letters were provided plus the surrounding text to help them generate the entire target word, which they then recorded. For targets presented intact, or the read targets, participants simply read and recorded the critical target word. In a following fill-in-the-blank test requiring participants to recall the missing critical targets given the rest of the sentence, participants performed significantly better for targets that had been generated versus those that had been read, thus demonstrating a generation advantage. Critically, however, when participants were then given a second passage to study, which once again consisted of a series of sentences containing critical words that were either presented intact or had to be generated from a fragment, a generation advantage was not observed on the same type of fill-in-the-blank test for the second passage. Importantly, the generation effect was eliminated, not because recall of the generated targets decreased, but because recall of the read targets increased. DeWinstanley and Bjork interpreted this finding as indicating that experiencing the generation advantage during the test following study of the first passage led learners to adopt more effective encoding strategies when faced with the task of learning the new information presented in the second passage.

Several studies have now both replicated the phenomenon reported by deWinstanley and Bjork and tested possible explanations of it (Bjork, deWinstanley, & Storm, 2007; Bjork & Storm, 2011; Burnett & Bodner, 2014). One possibility, explored by Bjork and Storm (2011), which they referred to as an enhanced contextual-processing account, involved the following assumptions. First, during study of the initial passage, participants used contextual information—specifically, the text surrounding the critical target word—to help them complete the generated targets. Then, during the fill-in-the-blank test, they became aware of the importance of having paid attention to contextual information when generating critical items during study for their being able to correctly recall the missing items for the fill-in-the-blank test questions. Consequently, during their study of the second passage, they employed this contextual-processing strategy when encoding critical items, regardless of whether those items were presented as to-be-generated or to-be-read items, thereby enhancing their performance for read targets to the level of that for generated targets.

To disambiguate the enhanced contextual-processing explanation of the results observed by deWinstanley and Bjork (2004) and by Bjork and Storm (2011), from accounts of the effects of generation on memory for intrinsic and extrinsic context—a related but quite different realm of inquiry in the literature (see, e.g., Mulligan, 2004, 2011; Marsh, Edelman, & Bower, 2001)—the enhanced contextual-processing account of Bjork and Storm can also be framed in the terms used by multifactor accounts of the generation effect (for relevant reviews, see Hunt & McDaniel, 1993; Mulligan & Lozito, 2004). According to such multifactor accounts, during the learning of a list of cue–target word pairs, in which some of the targets are to be generated and some are simply to be read, generation would act to enhance one or more of the following types of information, depending on the specific nature of the generation task: item-specific information (i.e., information specific to the target 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 particular generation task had primarily strengthened cue–target relational information, then on a later cued-recall test—assumed to be sensitive to such information (e.g., Einstein & Hunt, 1980; Tulving, 1962)—performance should be better for the generated cue–target pairs than for the read cue–target pairs. With regard to the materials and paradigm employed by deWinstanley and Bjork, their critical to-be-generated or to-be-read words can be thought of as the target word, and the surrounding context of the sentence can be thought of as the cue. Then, expressed in these terms, experiencing a generation advantage on the first test is thought

to lead participants to pay more attention to the processing of cue–target relational information when encoding the second passage, regardless of whether the target word is presented intact or with letters missing, thus potentially diminishing the size of the generation effect that would be expected on the test of the second passage.

Bjork and Storm (2011) tested this enhanced contextual-processing account by manipulating the type of test that participants experienced after studying the first passage. Specifically, instead of giving participants a fill-in-the-blank test, they were given a free-recall test requiring them to recall all the target words from the passage. Presumably, to perform well on such a free-recall test, participants would not have needed to encode the relationship between the target words and the surrounding context. Thus, although a generation advantage was observed on the free-recall test (as would be expected, given the greater strengthening of item-specific information that the generated items likely received), Bjork and Storm predicted that such an experience would be insufficient to induce learners to study the second passage in a way that would facilitate performance on read targets as they had done when receiving a fill-in-the-blank test (rather than a free-recall test) following study of the first passage. Consistent with this expectation, experiencing the benefits of generation on a free-recall test did not induce learners to study a second passage in a way that eliminated the generation advantage on a later fill-in-the-blank test. Bjork and Storm provided additional evidence for the enhanced contextual-processing account in a subsequent experiment by showing that participants who received a fill-in-the-blank test following the first passage remembered more surrounding context words (i.e., words other than the target words) from the second passage than did participants who received a free-recall test following the first passage.

Logic of the current research

The work of Bjork and Storm (2011) suggests that in order to become better learners of future to-be-read information, participants must experience an initial test that informs them as to the type of processing or encoding procedures that are going to be helpful on a subsequent test. It remains unclear, however, whether this test experience must also entail their becoming aware of a generation advantage. DeWinstanley and Bjork (2004) conjectured that participants must actually *experience* the benefits of generation on the test of the first passage—that is, gain metacognitive awareness of such benefits while taking the test—to realize how to study the second passage more effectively. This conjecture, which is consistent with the idea that people are unlikely to switch from a less effective learning strategy to a more effective learning strategy unless they actually experience the relative effectiveness of the two

strategies (e.g., Brigham & Pressley, 1988; Dunlosky & Hertzog, 2000; Shaughnessy, 1981), was supported by several further findings reported by deWinstanley and Bjork (Experiments 2 & 3). For example, they found the generation effect persisted on the test of the second passage—that is, participants failed to become better learners of read items in the second passage—when participants had the opportunity to experience only the encoding of one target type or the other (i.e., read or generate) on the first passage (cf. Burnett & Bodner, 2014).

It is possible, however, that neither experiencing a generation advantage nor even gaining an awareness of the generation advantage is necessary for participants to become better learners when studying the second passage. Rather, according to the most basic assumptions of the enhanced contextual-processing account, it would seem that for participants to become better learners of the second passage they would need only to become aware of (a) the type of criterion test that is going to be employed and (b) the type of information to which such a test is sensitive. In situations where the second test consists of fill-in-the-blank questions, for example, participants would need to discern that encoding target words in relation to the surrounding contextual words would be advantageous for performing well on such a test. Both of these criteria seem likely to be met without participants needing to experience, or become aware of, a generation advantage on the test of the first passage.

The two experiments of the present research were designed to help elucidate the necessary/sufficient conditions of the phenomenon in question by testing the contrasting predictions of the enhanced contextual-processing account (Bjork & Storm, 2011) versus the account proposed by deWinstanley and Bjork (2004). We did this by manipulating the type of experience that participants undergo during the learning of a first passage before being required to learn a second passage.

Experiment 1

In Experiment 1, participants were successively presented with two text passages that were to be learned, with each passage containing critical target words that either had to be generated or were presented intact and simply had to be read. For one set of participants, each passage was followed by a fill-in-the-blank test, in which participants had to retrieve the target words given the surrounding context words. This condition represented a direct replication of deWinstanley and Bjork (2004, Experiment 1), and thus we predicted that participants would become more effective encoders of read items in the second passage, possibly to the extent that the generation advantage would be eliminated or substantially reduced on the test of the second passage. For other participants, however, the fill-in-the-blank test for the first passage was

removed. Instead, these participants were either (a) given the opportunity to reflect on the differences between to-be-read and to-be-generated items from the first passage (self-reflect condition) or (b) given a brief description of the benefits of generation and, additionally, told that—had they been given a test—they would have recalled generate targets better than read targets (explanation-provided condition). Participants in both of these latter conditions were shown an example of the type of questions (i.e., fill-in-the-blank items) that they would have received had they been tested.

The self-reflect and explanation-provided conditions were designed to give participants the knowledge we speculated that participants typically possessed after experiencing the initial test (i.e., that generation led to a more effective strategy for encoding the critical targets than did simply reading them), but without actually giving them the opportunity to experience that test. If knowledge of a strategy's relative effectiveness is sufficient for learners to develop improved strategies for the future, then perhaps simply informing them of the relative effectiveness of each encoding strategy would be sufficient to lead participants to become better learners of the read items in the second passage. If so, then we would expect these participants to exhibit a pattern of results on the second passage similar to that exhibited by participants in the replication condition. If, however, participants actually need to experience an initial test, then participants in the explanation-provided and self-reflect conditions should show a very different pattern of results than participants in the replication condition. Specifically, they should exhibit diminished performance for target words presented as to-be-read items relative to those presented as to-be-generated items; that is, they should fail to exhibit an attenuation or elimination of the generation effect in their test performance for the second passage.

Method

Participants A total of 96 undergraduates at the University of Illinois at Chicago participated for credit in a psychology course. Participants were randomly assigned to the replication, self-reflect, and explanation-provided conditions.

Materials Text materials were two psychology-related passages (one on genetics, the other on self-esteem), and their order of presentation was counterbalanced across participants. Each passage was divided into individual sentences (8–23 words long), which were presented individually on a computer screen for 15 s each. The beginning of each passage contained two buffer sentences, which contained no critical target words. The next 14 sentences each contained a critical target word in red print presented in one of two ways: as an intact word (i.e., a read item) or as a word with missing letters (i.e., a generated item). For a read target, participants copied the word onto a lined sheet of paper in front of them. For a

generated target, participants tried to generate the word by filling in the missing letters and then writing the entire word onto the lined sheet of paper in front of them. Memory for the target words was assessed with fill-in-the-blank tests. The particular sentences containing read versus generated critical items was counterbalanced across participants.

Procedure All participants were tested individually. Participants were told they would be presented with a text passage, one sentence at a time, with each sentence after the first two containing one word appearing in red print. Participants were given a lined sheet of paper and asked to write down the red words they saw, by either filling in the missing letters for the fragmented words (generated targets) or simply copying the intact words (read targets). Immediately following study of the first passage, participants in the replication condition were given a fill-in-the-blank test, which consisted of sentences from the first passage as they had appeared on the computer screen, except with the critical target words missing, and they were given 2 min to recall as many of the missing critical targets as they could. Participants were not given any feedback with regard to their performance on this fill-in-the-blank test.

In contrast to the participants in the replication condition, who actually gained experience with what it was like to take a fill-in-the-blank test following study of the first passage, participants in the self-reflect and explanation-provided conditions were instead given a handout that asked them to imagine they had been given a fill-in-the-blank test for the information they had studied in the first passage. Participants in these conditions were also provided with an illustrative example of a type of question that would have appeared on the test (e.g., “Experimental psychology is a _____ essentially like any other. The answer is *science*.”). They were then asked if they thought they would be better able to remember the items they produced from fragments (generated targets) or the items presented intact (read targets). The majority (64 %) of participants indicated that they would remember the items they generated better than items they read.

After making their predictions, participants in the self-reflect condition were given two minutes to explain why they made their particular predictions. In contrast, participants in the explanation-provided condition were given two minutes to read a short paragraph (shown below) explaining the purported benefits of generation:

Memory researchers believe that you would have been better able to remember items that you produced from fragments. Having to generate an item from a fragment forced you to learn that item in the context of the other words in the sentence. Learning how each item fits within each sentence would have helped you perform better on the fill-in-the-blank test.

They were then asked to indicate how much they agreed with this reasoning on a 9-point Likert scale, with 1 representing *not at all*, 5 corresponding with *unsure*, and 9 representing *very much*. Participants gave a median rating of 7.5, with 94 % of participants giving a rating of 6 or above. Thus, in general, participants in the explanation-provided condition seemed to have come to an appreciation of the potential benefits of generation.

Participants in all three conditions were then presented with a second passage that once again contained to-be-read and to-be-generated target words, followed immediately by the same type of fill-in-the-blank test administered to participants in the replication condition after their study of the first passage.

Results and discussion

Targets were only marked as correct if they were recalled in the appropriate sentence. Minor spelling errors (e.g., “intamacy”) were marked as correct.

The top panel of Table 1 shows the mean correct recall proportions for the critical target items from each passage and condition. The replication group’s recall performance was analyzed using a 2 (target type: generated vs. read) \times 2 (passage: first vs. second) repeated-measures analysis of variance (ANOVA). Significant main effects were observed such that recall was higher for the generated targets ($M = .51$, $SE = .04$) than for the read targets ($M = .40$, $SE = .03$), $F(1, 31) = 12.59$, $MSE = .03$, $p = .001$, partial $\eta^2 = .29$, and overall recall increased from the first passage ($M = .35$, $SE = .03$) to the second passage ($M = .56$, $SE = .04$), $F(1, 31) = 35.08$, $MSE = .04$, $p < .001$, partial $\eta^2 = .53$. Importantly, the interaction was also significant, $F(1, 31) = 11.03$, $MSE = .02$, $p = .002$, partial $\eta^2 = .26$. For the first passage, generated targets ($M = .45$, $SE = .05$) were recalled significantly better than the read targets ($M = .26$, $SE = .03$), thus exhibiting a standard generation

advantage, $t(31) = 4.43$, $p < .001$, $d = .81$. For the second passage, however, generated targets ($M = .58$, $SE = .05$) were not recalled significantly better than the read targets ($M = .54$, $SE = .04$), $t(31) = 1.02$, $p = .32$, $d = .18$. Although the numerical advantage for generated targets was not completely eliminated across passages, recall significantly improved for read targets, a result that is consistent with that of deWinstanley and Bjork (2004, Experiment 1).

Next, to assess whether participants actually need to experience an initial test to develop improved encoding strategies, we examined whether participants in the other two conditions exhibited a significant generation advantage in their recall of critical targets presented in the second passage. In the self-reflect condition, we found participants to recall significantly more generated targets ($M = .49$, $SE = .05$) than read targets ($M = .36$, $SE = .03$), $t(31) = 3.53$, $p = .001$, $d = .62$. Similarly, in the explanation-provided condition, we found participants to recall significantly more of the generated targets ($M = .36$, $SE = .04$) than read targets ($M = .27$, $SE = .04$), $t(31) = 2.14$, $p = .04$, $d = .38$. Thus, a generation advantage still emerged in the recall of these two groups, even though both groups had been given an initial study experience with read and generated targets and had seemed either to understand that generation should be a more effective encoding strategy (i.e., the self-reflect participants) or had been explicitly told that generation would be the more effective encoding strategy (i.e., the explanation-provided group).

To compare recall performance in the three conditions on the test of the second passage, we conducted a 2 (target type: generated vs. read) \times 3 (condition: replication vs. self-reflect vs. explanation-provided) mixed-design ANOVA. Significant main effects of both target type, $F(1, 93) = 15.06$, $MSE = .02$, $p < .001$, partial $\eta^2 = .14$, and condition, $F(2, 93) = 10.91$, $MSE = .09$, $p < .001$, partial $\eta^2 = .19$, were observed. Despite the apparent pattern illustrated in Table 1, however, the interaction effect was not significant, $F(2, 93) = 1.57$, $MSE = .02$, $p = .21$, partial $\eta^2 = .03$, advocating that caution should be taken in the interpretation of these data. Although we can be confident in our conclusion that participants in the replication condition exhibited a significantly diminished generation effect in their performance on the tests of the second versus the first passage, we cannot conclude that this diminished effect is different from that observed in the other two conditions. It is interesting to note, however, that the significant generation effects revealed in the previous analyses for both the self-reflect and explanation-provided conditions are numerically smaller (13 % and 9 %, respectively) than that observed in the test of the first passage in the replication condition (19 %), suggesting that the self-reflect and explanation-provided conditions may have been at least somewhat effective in reducing the generation advantage. It is also possible that not all participants in the explanation-provided condition fully understood the explanations that were provided, in which case we might

Table 1 Mean Correct Recall Proportions for Critical Targets as a Function of Condition in Experiments 1 and 2 (Standard Errors in Parentheses)

| Experiment | Passage 1 | | Passage 2 | |
|----------------------|-----------|-----------|-----------|-----------|
| | Generate | Read | Generate | Read |
| Experiment 1 | | | | |
| Replication/Test | .45 (.05) | .26 (.03) | .58 (.05) | .54 (.04) |
| Self-Reflect | – | – | .49 (.05) | .36 (.03) |
| Explanation-Provided | – | – | .36 (.04) | .27 (.04) |
| Experiment 2 | | | | |
| Generate-Advantage | .67 (.04) | .26 (.04) | .62 (.03) | .56 (.04) |
| Read-Advantage | .44 (.04) | .60 (.04) | .61 (.03) | .59 (.04) |
| Overall | .56 (.03) | .43 (.03) | .62 (.02) | .57 (.03) |

expect the generation effect to be only partially diminished and not completely eliminated.

We next followed up on the significant main effect of condition by conducting three independent-sample comparisons of overall recall performance in the test of the second passage. All three comparisons were statistically significant. Specifically, participants in the replication condition ($M = .56$, $SE = .04$) outperformed participants in both the self-reflect condition ($M = .42$, $SE = .04$), $t(62) = 2.44$, $p = .02$, $d = .61$, and explanation-provided condition ($M = .31$, $SE = .03$), $t(62) = 4.57$, $p < .001$, $d = 1.14$; and participants in the self-reflect condition outperformed participants in the explanation-provided condition, $t(62) = 2.27$, $p = .03$, $d = .57$.

The finding that overall performance was better in the replication condition than in the other two conditions seems quite reasonable. By actually experiencing the first test, these participants would likely have been more prepared to do well on the second test, or put in terms of the generation accounts under consideration, would have become more aware of the type of information to which such tests are sensitive. The reason for the difference in performance between the self-reflect and explanation-provided conditions is, however, a bit more speculative. One possibility is that thinking about the effects of generation after study of an initial passage may be more effective for enhancing subsequent encoding than being told about a strategy that would be beneficial for such encoding. Perhaps the self-reflect instructions led participants to think more carefully about the nature or demands of a fill-in-the-blank test and to speculate whether their performance on such a test might be better for targets they had needed to produce from fragments or for ones that had been presented to them intact. Such reasoning may have then led these participants to realize more fully the type of processing in which they should engage during future study in order to do well on such a test than could participants who were simply provided an explanation. Although the precise reason for the difference in performance between the two conditions is unclear, what is clear (at least in this context) is that being told what to do while studying can be less effective than being allowed to self-reflect on what one should do while studying.

Experiment 2

The pattern of results obtained in Experiment 1 suggests that simply being told about the benefits of generation may not be sufficient for participants to become better encoders of read items when learning a subsequent passage, at least to the extent that read items are learned as well as generated items. Rather, to achieve this latter outcome, it appears that participants need actually to experience the benefits of generation themselves in the context of a test. At first glance, this finding appears to be consistent with deWinstanley and Bjork's

(2004) assumption that participants need to experience better performance for generated items than read items on the initial test—and that it is this experience that then drives participants to engage in more effective encoding procedures for read items on the second passage. It is also possible, however, that such specific awareness—or the actual experiencing of the generation advantage in the context of a test—is unnecessary. Perhaps all that is necessary is experience with, or even just information about, the nature of the generation task combined with the experience of taking the type of test that will be employed in the future. Perhaps access to the knowledge that comes with these experiences would be sufficient to allow participants to realize how the nature or demands of the generation task might have led them to encode information during study of the first passage in a way that then enhanced their later test performance. If so, under such circumstances, participants might be able to develop more effective encoding strategies for processing subsequent to-be-read information, even if they had not actually experienced a memorial advantage of generation on the test of the first passage.

Although it seems clear from both the present and previous results that participants learn something from the testing of the first passage that leads them to become better learners of the second passage, it may not be their relative levels of recall performance for generate versus read items that drive this change. Rather, it may be the experience of performing both types of tasks plus the taking of the test that allows them to realize how their encoding of information during the generation task versus the reading task might affect their later performance on such a test. If so, then perhaps it would not matter whether participants do or do not actually experience a memorial advantage for generation in the context of a test for them to improve their encoding strategies for the processing of future information. Although this question has been addressed in a study by Burnett and Bodner (2014), a limitation of their study was that the extent to which participants exhibited a generation advantage on the first passage was not experimentally manipulated. Instead, participants were split into two groups based on whether their performance happened to exhibit a read advantage or a generate advantage, thus leaving the results open to being influenced by selection effects and/or differences in counterbalancing.

In the present Experiment 2, we actually manipulated whether participants would or would not experience a benefit of generation on the test of a first passage, with a random half of the participants experiencing a benefit of generation on the first test (generate-advantage condition) and the other random half experiencing a benefit of reading on the first test (read-advantage condition). If participants need to actually experience a generation advantage in their own test performance to become better learners of future to-be-read information, then participants in the generate-advantage condition should exhibit a greater increase in recall for to-be-read targets on the

second passage than participants in the read-advantage condition.

Method

Participants Eighty undergraduates (M age = 20.4) from the University of California, Santa Cruz (UCSC), participated for credit in a psychology course. The data from one participant was removed due to experimenter error.

Materials and procedure The materials and procedure were nearly identical to those used in the replication condition of Experiment 1. Participants were presented with two passages, each containing a series of sentences with a critical target word that was either to be read or to be generated. Unbeknownst to the participants, however, the first passage was constructed to make either generate or read targets “advantaged” and thus easier to recall on the later test. Using data from Experiment 1, we identified the seven target words that were easiest to recall (average recall: 58 %) and the seven target words that were most difficult to recall (average recall: 34 %), regardless of encoding condition. Then, for participants in the generate-advantage condition, the to-be-generated targets all came from the easy set of items while the to-be-read targets all came from the difficult set of items. In contrast, for participants in the read-advantage condition, the to-be-generated targets all came from the difficult set of items and the to-be-read targets all came from the easy set of items. For both groups of participants, however, the second passage presented for study was constructed in the same way as the second passage in the replication condition of Experiment 1; specifically, the items were randomly distributed as read or generated targets and counterbalanced.

Results and discussion

We first conducted a 2 (target type generated vs. read) \times 2 (passage: first vs. second) \times 2 (condition: generate-advantage vs. read-advantage) mixed-design ANOVA with the first two variables serving as repeated measures. Significant main effects were observed such that overall recall was higher for the generated targets ($M = .59$, $SE = .02$) than for the read targets ($M = .50$, $SE = .02$), $F(1, 77) = 17.23$, $MSE = .03$, $p < .001$, partial $\eta^2 = .11$, and overall recall increased from the first passage ($M = .49$, $SE = .02$) to the second passage ($M = .60$, $SE = .02$), $F(1, 77) = 19.54$, $MSE = .04$, $p < .001$, partial $\eta^2 = .20$. More importantly, replicating the work by deWinstanley and Bjork (2004), the interaction between target type and passage was significant, $F(1, 77) = 5.41$, $MSE = .02$, $p = .02$, partial $\eta^2 = .04$. For the first passage, the generated targets ($M = .56$, $SE = .03$) were recalled significantly better than the read targets ($M = .43$, $SE = .02$), thus exhibiting a generation advantage, $t(78) = 2.89$, $p = .005$, $d =$

.32. For the second passage, however, although the numerical advantage for generated targets over read targets was not completely eliminated, generated targets ($M = .62$, $SE = .02$) were not recalled significantly better than read targets ($M = .57$, $SE = .03$), $t(78) = 1.57$, $p = .12$, $d = .18$. Finally, as expected, a significant three-way interaction was observed, $F(1, 77) = 57.47$, $MSE = .02$, $p = .001$, partial $\eta^2 = .43$. To gain a deeper understanding of the pattern of overall results, we next examined performance on the first and second passages separately, as described below.

First-passage recall performance Performance on the first passage was analyzed to confirm that a memorial advantage for generated targets had been obtained in the generate-advantage condition and that a memorial advantage for read targets had been obtained in the read-advantage condition. As can be seen on the left side of the bottom two rows of Table 1, and as confirmed by a 2 (condition: generate-advantage vs. read-advantage) \times 2 (target type: generated vs. read) ANOVA, this interaction is exactly what we observed, $F(1, 77) = 120.51$, $MSE = .03$, $p < .001$, partial $\eta^2 = .61$. Specifically, participants in the generate-advantage condition recalled significantly more generated targets ($M = .67$, $SE = .04$) than read targets ($M = .26$, $SE = .04$), $t(38) = 11.18$, $p < .001$, $d = 1.79$, whereas participants in the read-advantage condition recalled significantly more read targets ($M = .60$, $SE = .04$) than generated targets ($M = .44$, $SE = .04$), $t(39) = 4.39$, $p < .001$, $d = .69$. In fact, whereas 36 of the 39 participants in the generate-advantage condition recalled generated targets better than read targets, only 5 of the 40 participants in the read-advantage condition recalled generated targets better than read targets. All data were included in the analyses below. It should be noted, however, that the same pattern of results was observed when participants who did not exhibit the appropriate difference in performance were removed.

Second-passage recall performance We next analyzed recall performance for critical target words in the second passage to see whether participants in the generate-advantage condition exhibited a different pattern of results than participants in the read-advantage condition. As can be seen on the right side of Table 1, and as confirmed by a 2 (condition: generate-advantage vs. read-advantage) \times 2 (target type: generated vs. read) ANOVA, they did not. The main effect of condition was not significant, $F(1, 77) = .08$, $MSE = .06$, $p = .78$, partial $\eta^2 = .00$, nor was the main effect of target type, $F(1, 77) = 2.50$, $MSE = .03$, $p = .12$, partial $\eta^2 = .03$, or the interaction between condition and target type, $F(1, 77) = .70$, $MSE = .03$, $p = .40$, partial $\eta^2 = .01$. Although not significant differences, it is interesting to note that the generation advantage was numerically greater (and read items were recalled numerically less often) in the generate-advantage condition than in the read-advantage condition, which is exactly opposite of what would

be expected based on the idea that participants need to experience an advantage for generated items over read items on the test of the first passage to then alter the way in which they encode the to-be-read items on the second passage. These results suggest that the extent to which participants experience a generation advantage on the test of an initial passage does not influence the extent to which the effect is then eliminated on the test of a second passage.

Replication of read-advantage condition with feedback on performance in first passage One limitation of Experiment 2 is that it is unclear whether participants were actually aware of their relative performance vis-à-vis read versus generate items on the test of the first passage. It is possible, for example, that although participants in the read-advantage condition failed to exhibit a generation effect, they may have still subjectively perceived themselves as performing better on generate items than on read items. (Such an illusion would be similar to that observed in studies comparing the effects of interleaving schedules versus blocked schedules in the learning of related skills or concepts, which have shown that participants, even after performing better following interleaved practice tend to believe that they learned better under blocked practice, e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kornell & Bjork, 2008; Simon & Bjork, 2001).

To address this issue, we conducted a replication of the read-advantage condition ($n = 40$ UCSC undergraduates) with one important change to the methodology. Specifically, after participants finished the test of the first passage, the experimenter went through their test with them, scoring the items and identifying whether each item had been generated or read while learning the passage. The experimenter then counted up the number of items that the participant answered correctly in each condition, thus informing each participant of whether they, in fact, had or had not exhibited a generation effect. The 26 participants who failed to exhibit a generation effect on the first passage (and were made aware of this fact) did not exhibit a generation effect on the second passage, $t(25) = .31$, $p = .76$, $d = .06$ (Generate: $M = .66$, $SE = .05$; Read: $M = .64$, $SE = .05$). As expected, a nonsignificant effect was also observed when data from all 40 participants were analyzed, $t(39) = .36$, $p = .72$, $d = .06$ (Generate: $M = .67$, $SE = .03$; Read: $M = .68$, $SE = .04$). Based on these results, it seems clear that even when participants are explicitly made aware that they had not exhibited a generation advantage in their performance on the first test, they are still able to engage in an encoding strategy during study of the second passage that brings their recall of read and generated targets to the same level of performance on the test of that passage.

General discussion

The present research demonstrates that learners, under certain conditions, can improve their encoding of new information without explicit instructions on how to do so. In Experiment 1, we replicated a phenomenon observed by deWinstanley and Bjork (2004), in which participants who were made sensitive to the memorial benefits of generation in the context of a test following study of a passage in which they had encoded critical words via both generation and reading then became more effective learners of future to-be-read information presented in a second passage. Just as in this previous work by deWinstanley and Bjork, participants in our replication condition improved their encoding of read targets enough to raise their recall of them significantly and to a level that did not significantly differ from that for the generated targets. We did not, however, find evidence in support of deWinstanley and Bjork's suggestion that actually experiencing the memorial advantage of generation in the context of a test is necessary for the development of an improved encoding strategy for future to-be-read information. Rather, our pattern of results seems more consistent with the enhanced contextual-processing account suggested by Bjork and Storm (2011) and, more generally, with test-expectancy findings, such as those of Finley and Benjamin (2012). Namely, that for new information to be encoded with a more effective processing strategy, learners need to be made aware of the type of test they will be receiving and the type of information to which such a test is sensitive—that is, the type of information one needs to encode during study to perform well on such a test. In the case of the Bjork and Storm study, experience with a fill-in-the-blank test seemed to inform participants about the advantages of paying attention to contextual information during encoding of the sentences, something that participants had probably done (rather automatically) when required to generate targets but not when only required to read them during study of an initial passage. Equipped with such knowledge after the test experience, however, they then appeared to pay attention to surrounding contextual information during encoding of the second passage for both the to-be-generated targets and the to-be-read targets.

Although in the present Experiment 1, it was not the case that participants were simply told about or asked to reflect upon the possible benefits of generation for their later performance on a fill-in-the-blank test and then went on to eliminate a generation advantage in the test of a second passage, other aspects of our results nonetheless suggest that actually experiencing the generation benefit in the context of a test is not a necessary condition for improving encoding of future to-be-read information. First, participants in the self-reflect condition of Experiment 1, who reflected on the possible benefit of generation for performance on a fill-in-the-blank test, but who did not actually experience such a benefit, did produce

significantly higher overall recall for the second passage than did participants who were just told about the benefit of generation (the explanation-provided condition). We see this difference in the test performance of the two groups as indicating that participants in the self-reflect condition, as compared to those in the explanation-provided condition, were led to think more carefully about the nature or demands of a fill-in-the-blank test in relation to the generation task versus the read task that they had performed during study of the first passage, which, in turn, may have helped them to imagine more accurately (or at least make them more likely to employ) the type of processing in which they should engage during future study in order to do well on such a test. Second, in Experiment 2, we found that participants could improve their study strategy on a second passage regardless of whether their recall performance on a fill-in-the-blank test of the first passage exhibited a generation advantage. Taken together, then, the results of the present experiments suggest that—in contrast to the assumption of deWinstanley and Bjork (2004)—it is not necessary for learners to actually experience the memorial advantage of generation to change how they process future to-be-read information.

Still, given that participants in the self-reflect and the explanation-provided conditions of Experiment 1, unlike the replication-condition participants, continued to show a generation advantage on the test for the second passage, it would seem to be the case that the test experience can play an important role in prompting an effective strategy switch. Perhaps participants who take an actual test gain a clearer picture of its demands and the type of information to which it is sensitive, with this greater knowledge enabling them to make appropriate qualitative changes in the processing strategy they use during study of a second passage to a greater extent than can participants not having an actual test experience. Given the lack of a significant interaction among the three conditions, however, we think this interpretation must be made with caution—that is, although the test experience may be helpful in promoting a switch in processing strategy, we cannot argue that such an experience is necessary. Indeed, in contrast to the various findings indicating the importance of a testing experience or knowledge about the demands of an upcoming test for improving processing strategies, Burnet and Bodner (2014) have reported a study in which the generation advantage for the second passage was eliminated in a condition where participants were not given a test for the first passage.

On the basis of the overall pattern of results observed across the present studies, and those from the test-expectancy literature (e.g., Finley & Benjamin), we suggest that two different factors play important roles in leading participants to develop improved encoding strategies in their future study: (a) the opportunity to have some experience with or knowledge about different types of encoding strategies (e.g., generating vs. reading in the present studies) and (b)

knowledge about the nature of the test they are to be given on the material they are being asked to learn, with this knowledge gained either through the actual experience of taking such a test or in some other way. Furthermore, while the presence of either of these factors would perhaps be sufficient for participants to improve their future encoding strategies somewhat, the presence of both of them would lead to a greater amount of improvement.

These assumptions would be consistent with the general pattern of results obtained in the present research as well as those obtained in the studies of deWinstanley and Bjork (2004), Bjork and Storm (2011), and Burnett and Bodner (2014). Moreover, the differences in the results obtained in deWinstanley and Bjork's Experiments 1A and 1B and those obtained in their Experiment 3 would seem to illustrate the need for both of these factors to be in operation to gain the greatest amount of improvement to the processing of information presented in the second passage—that is, to raise the encoding and recall of read information to that of generated information. In their Experiments 1A and 1B, where both factors were in place, a generation advantage was eliminated in the participants' performance on the test of the second passage. In their Experiment 3, however, whereas all participants learned about the nature of the test by taking it after study of the first passage, because participants only generated targets or only read targets during study of the first passage, none of them had the opportunity to gain some experience with or knowledge about the differences between the types of encoding processes potentially invoked by the generation task versus the read task. Thus, while participants in both the read and generate groups of their Experiment 3 were able to raise their performance across passages, which could be interpreted as reflecting a general learning-to-learn type of benefit, the performance of participants in the read group did not rise to the level of the generate group in the test of the second passage.

Finally, the potential application of the findings of the present research to educational practice seems worth mentioning. Prior research has shown that metacognitive judgments of learning are often misguided and subject to illusions and that, left to their own devices, students do not often engage in optimally effective learning strategies (e.g., Bjork, Dunlosky, & Kornell, 2013; Koriat, 1997). Might learning about the nature and/or demands of upcoming tests and the types of information to which they are sensitive be a more promising way to lead students to adopt more effective strategies for acquiring and remembering information? The use of practice tests, for example, may be a particularly effective way to help students more appropriately prepare for upcoming examinations. Practice tests using free-recall, cued-recall, and competitive multiple-choice formats have all been shown to increase long-term retention of the tested material (see, e.g., Bjork, Little, & Storm, 2014; Roediger & Karpicke, 2006a, b;

Roediger, Putnam, & Smith, 2011), as well as to improve metacognitive control of future study-time allocation (see, e.g., Kornell & Bjork, 2007; Little & McDaniel, 2015). Experience with practice tests also provides students with the chance to learn the format of future exams and to adjust their study strategies accordingly. As Finley and Benjamin (2012) argued, studying is most efficient when learners are equipped with beneficial encoding strategies.

Although learning in educational settings is often more complex and varied than that exemplified by the tasks of the present research, the ability of learners in the present research to anticipate the demands of an upcoming test and to select more appropriate encoding strategies as a consequence of that experience seems promising for improving learning both within the classroom and in contexts involving self-regulated learning. Considering that only a small minority of students report ever being taught study strategies (Kornell & Bjork, 2007), it is encouraging to know that learners can discover more effective encoding strategies for themselves through experience. Indeed, it would seem to be the case that some encoding strategies are best learned through active participation and practice, with learners being afforded the opportunity to become familiar with the demands of a test or task by personally engaging with it, as opposed to being told by educators the specific ways in which they should study.

Acknowledgments The authors thank Sanaz Batmanghilih, Jamie Grinkevich, Sean Kotzman, Amelie Meltzer, Prachi Mistry, Payton Small, and Nathalie Zhen for their assistance with data collection and coding.

References

- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*, 201–210. doi:10.3758/BF03193441
- Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition, 41*, 392–402. doi:10.3758/s13421-012-0272-7
- 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., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444. doi:10.1146/annurev-psych-113011-143823
- Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition, 3*, 165–170. doi:10.1016/j.jarmac.2014.03.002
- Bjork, E. L., & Storm, B. C. (2011). Retrieval experience as a modifier of future encoding: Another test effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1113–1124. doi:10.1037/a0023549
- Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3*, 249–257.
- Burnett, A. N., & Bodner, G. E. (2014). Learnin’ ‘bout my generation? Evaluating the effects of generation on encoding, recall, and metamemory across study-test experiences. *Journal of Memory and Language, 75*, 1–13. doi:10.1016/j.jml.2014.04.005
- 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. (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. (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*, 588–589. doi:10.1037/0278-7393.6.5.588
- Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38*, 632–652. doi:10.1037/a0026215
- 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
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421–445. doi:10.1006/jmla.1993.1023
- 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
- 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
- Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219–224. doi:10.3758/BF03194055
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19*, 585–592. doi:10.1111/j.14679280.2008.02127.x
- Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition, 43*, 85–98. doi:10.3758/s13421-014-0453-7
- Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstration of a generation effect in context memory. *Memory & Cognition, 29*, 798–805. doi:10.3758/BF03196409
- McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (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: 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
- Metcalf, 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
- Metcalf, 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
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. doi:10.1016/S0022-5371(77)80016-9
- Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 838–855. doi:10.1037/0278-7393.30.4.838
- Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not extrinsic context. *The Quarterly Journal of Experimental Psychology*, *64*, 1543–1562. doi:10.1080/17470218.2011.562980
- Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 45, pp. 175–214). San Diego: Academic Press.
- 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., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x
- Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives of Psychological Science*, *1*, 181–209. doi:10.1111/j.1745-6916.2006.00012.x
- Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford: Elsevier.
- 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
- Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review*, *69*, 344–354. doi:10.1037/h0043150