Learnin’ ‘bout my generation? Evaluating the effects of generation on encoding, recall, and metamemory across study-test experiences

A. Nicole Burnett and Glen E. Bodner

University of Calgary

In press, Journal of Memory and Language [accepted April 28, 2014]

Correspondence to: Glen E. Bodner
Department of Psychology
University of Calgary
2500 University Drive NW
Calgary AB T2N 1N4
CANADA

phone: 403-220-2714
fax: 403-282-8249
e-mail: bodner@ucalgary.ca
Abstract

We explored how learning during an initial study-test experience with text materials shapes future encoding, recall, and metamemory. Differential recall of targets from generate and read sentences on a fill-in-the-blank test led participants to shift their encoding strategies such that differential recall was eliminated on a second study-test block using different materials. This shift was not contingent on experiencing a generation advantage on the first test: recall also improved across tests when groups studied and recalled only one target type, did not receive the initial test, or showed a null or negative generation effect on the initial test. Strategy reports suggest that a sentence-target linking strategy increased across tests. Importantly, metamemory measures failed to reveal awareness of differential performance for read and generate targets. Contrary to recent claims, then, our findings suggest that individuals can learn, perhaps even tacitly, to modify their study strategies based on an initial study experience.

Keywords: learning; memory; recall; metamemory; generation; strategies
Learnin’ ‘bout my generation? Evaluating the effects of generation on encoding, recall, and metamemory across study-test experiences

People have a multitude of strategies at their disposal for committing new information to memory, from rote repetition to mnemonics. Memory research has identified many effective study strategies such as elaboration (Craik & Lockhart, 1972), generation (Slamecka & Graf, 1978), and simple production (Macleod, Gopie, Hourihan, Neary, & Ozubko, 2010). Although there has been extensive research on the effects of such strategies, there has been relatively little examination of their influence on future encoding and retrieval simply because most memory studies consist of a single study phase followed by a single memory test. More than one study-test block is needed to gauge what people learn from using a study strategy and how that experience modifies subsequent encoding and memory.

By using multiple study-test blocks, researchers can reveal how experience leads people to adapt their approaches to learning new information. Indeed, a number of important memory effects have been identified using multi-trial paradigms including the testing effect (e.g., Roediger & Karpicke, 2006), spaced retrieval-practice effects (e.g., Whitten & Bjork, 1977), retrieval-induced facilitation (e.g., Chan, 2009), and the practice effect (e.g., Postman, Burns, & Hasher, 1970). Research on metamemory—how people think about and monitor their encoding and retrieval processes—has also employed multiple study-test designs (e.g., Brigham & Pressley, 1988; Dunlosky & Hertzog, 2000; Hertzog, Price, & Dunlosky, 2008; Hertzog et al., 2009; Koriat, 1997; Tiede & Leboe, 2009).

Dunlosky and Hertzog (2000; see also Hertzog et al., 2008; 2009) developed a detailed metacognitive framework of strategy knowledge updating across study-test experiences. Their knowledge-updating framework assumes that study strategies vary in their effectiveness for
enhancing memory (the *effectiveness assumption*), and individuals monitor their study and/or test experiences and use them to choose/update study strategies (the *utilization assumption*). These monitoring and evaluation processes yield knowledge updating if two conditions are met. First, individuals must monitor their strategies as they study and/or monitor their performance as their memory is tested (the *monitoring assumption*). Second, individuals must attribute their memory performance to the study strategies they used and then update their knowledge about those strategies (the *updating assumption*). By collecting both metamemory and memory measures at various points across two study-test blocks, and then using path analyses to chart the links between these measures, these researchers established a very useful, testable model. However, the encoding strategies they examined across study-test experiences was rote repetition versus imagery, and importantly, although metamemory measures shifted across study-test experiences, the advantage of imagery over repetition did not.

We examined a different study strategy in the present work, namely *generation*, because de Winstanley & Bjork (2004; see also Bjork, de Winstanley, & Storm, 2007; Bjork & Storm, 2011; Bjork, Storm, & de Winstanley, 2011) revealed a striking shift in its effect on recall across study-test experiences. The generation effect refers to a typically robust memory advantage from self-generation of a target (e.g., k_tt_n) relative to simple reading (e.g., kitten; Hirshman & Bjork, 1988; McDaniel, Waddill, & Einstein, 1988; Slamecka & Graf, 1978; see Bertsch, Pesta, Wiscott, & McDaniel, 2007, for a review). Effective generation tasks include answering questions, solving anagrams, or, as was the case here, solving word fragments.

de Winstanley and Bjork’s (2004) participants studied a paragraph on a particular topic. Each sentence contained a target in red. Half the targets were intact and were read silently (*read* targets), and half the targets had to be generated from word fragments (*generate* targets).
Memory for the targets was then tested using a fill-in-the-blank test where the same sentences were presented with the targets left blank. This Block 1 procedure was then repeated in Block 2 using a different paragraph on a different topic.

de Winstanley and Bjork’s (2004) key result was that a generation effect occurred on Test 1 but not on Test 2. The elimination of the generation effect on Test 2 was attributed to improved recall of read targets across tests. de Winstanley and Bjork argued that participants experienced the relative benefits of generation during Test 1 (i.e., monitoring as per Dunlosky & Hertzog, 2000), that lead them to develop an improved study strategy for the read targets during Study 2 (i.e., updating and utilization as per Dunlosky & Hertzog). As stated in the article title, their participants thus appear to have become “better readers” after experiencing the generation effect. Importantly, the generation effect persisted (albeit a between-subject generation effect) when participants received only one target type per block or across blocks, suggesting that directly experiencing the generation effect on Test 1 was critical to this strategy knowledge updating.

Using the same paradigm, Bjork and Storm (2011, Experiment 3) identified an important boundary condition (for another boundary condition, see Burnett, 2013): the generation effect persisted on Test 2 when Test 1 was a free recall test. The researchers argued that unlike a fill-in-the-blank test, free recall did not enable participants to learn that linking each target word with its sentence (what we term a context strategy) could benefit memory, thus they did not modify their study strategy for read targets on Block 2.

Bjork and Storm (2011, Experiment 4) provided further evidence that participants adopted a context strategy for Block 2. The design was the same as Experiment 3, except Test 2 tested participants’ memory for a word from each sentence context rather than for the target itself. Participants recalled more context words on Test 2 when Test 1 was a fill-in-the-blank test.
rather than free recall. Thus, exposure to a fill-in-the-blank test may have led participants to shift to a context strategy for Block 2. Bjork and Storm suggested that this strategy shift eliminated the generation effect on Test 2. Strategy reports provided some support for this possibility. However, Bjork and Storm did not provide direct evidence that use of a context strategy increased *target* recall on Test 2.

In sum, de Winstanley and Bjork (2004) posited that experiencing the benefits of a study strategy on an initial test (a monitoring effect) could spawn shifts in encoding strategies (an updating effect) that influenced subsequent memory performance (a utilization effect). We report three experiments using their paradigm that provide a detailed evaluation of this claim. In Experiment 1, we established a replication of de Winstanley and Bjork’s (2004, Experiment 1) within group, and we then compared shifts in recall of read and generate targets in this group to pure-list read and generate groups who did not experience the generation effect. The increase in recall of read targets across tests should be larger in the within group than the read group if experiencing the generation effect is critical to the boost in read target recall on Test 2. de Winstanley and Bjork (2004) tested their within and between groups in separate experiments using different stimuli and thus were unable to evaluate these possibilities. Participants’ self-reported study strategies were also collected in Experiment 1 (see Hertzog et al., 2008; 2009) to determine what study strategies people use, whether they shift strategies across blocks, and to evaluate whether the within group shifted to a context strategy selectively for the read targets.

Experiment 2 tested de Winstanley and Bjork’s (2004) claim that participants become aware of the relative benefits of generation during Test 1. To this end, a within group was compared to a second within who did not receive a Test 1. If participants learn about the relative
benefits of generation during Test 1, then only the within group who receives Test 1 should show the elimination of the generation effect on Test 2.

Finally, Experiment 3 collected several metamemory judgments (after Dunlosky & Hertzog, 2000) to determine whether and when participants become sensitive to the generation strategy. We also examined whether shifts in metamemory judgments were concordant with shifts in recall. Bjork and Storm (2011) found that participants’ metamemorial reports of context strategy use were predictive of their recall pattern. If participants learn about the benefits of generation during their initial study experience, they should provide higher metamemory judgments for generate targets prior to Test 1. In contrast, if participants learn about the benefits of a study strategy primarily during a test, as suggested by both de Winstanley and Bjork (2004) and Hertzog et al. (2008; 2009), then only metamemory judgments collected during/after Test 1 (cf. during/after Study 1) should differ for read and generate targets.

**Experiment 1**

Experiment 1 tested whether experiencing a generation manipulation on an initial test is critical to the elimination of the generation effect on a second test. A *within group* performed two study-test blocks in which different sets of read and generate sentences appeared on each study list and on each fill-in-the-blank test. This group was expected to replicate de Winstanley and Bjork (2004) by showing a generation effect on Test 1 and its elimination on Test 2 due to improved recall of read targets. Experiment 1 also included two pure-list *between groups*. In the *read group*, only read sentences were studied and tested in each study-test block; this group examined whether recall for read targets improves across tests when the generation manipulation is not experienced. In the *generate group*, only generate sentences were studied and tested in each study-test block; this group was used to examine the possibility that memory for generate
targets in the within group suffers a hidden cost across tests. Although de Winstanley and Bjork found equivalent recall for generate targets across tests in a within group, recall might increase across tests in a pure generate group. For example, recall of generate targets in the within group might not improve across tests if participants shift their encoding strategies toward the read sentences during Study 2. de Winstanley and Bjork (2004) also tested both within and pure-list groups, but they did so in separate experiments using different materials, and they did not directly compare these groups’ recall.

Half of each group completed a post-experiment questionnaire to identify their self-reported study strategies for each paragraph, and approach used in related work (Hertzog et al., 2008; 2009). This questionnaire also allowed us to examine Bjork and Storm’s (2011) claim that the elimination of the generation effect on Test 2 is due to an increased linking of read targets with their sentences during Study 2 (i.e., a context strategy).

**Method**

**Participants**

University of Calgary undergraduates enrolled in at least one psychology course participated for course credit. They were randomly assigned to the within, read, or generate group (n = 96 each). Experiment 1 averages across a successful replication where a post-experiment questionnaire was added (n = 48 each). This pool was also drawn from, without replacement, for Experiments 2 and 3.

**Materials**

The materials were the two psychology-related paragraphs (one on motivation and goal orientation; one on Bloom’s taxonomy of instructional objectives) used in de Winstanley and Bjork (2004, Experiment 1B) and Bjork and Storm (2011). Each paragraph contained 2 buffer
sentences then 10 critical sentences. Critical sentences each contained a target word in red (underlined in the forthcoming examples) that was studied one of two ways. The *read* sentences were read silently (e.g., “The emotional or *affective* part.”). For *generate* sentences, the red target appeared as a word fragment for participants to solve (e.g., “The emotional or *aff-ct-v-* part.”). For the fill-in-the-blank test the 10 critical sentences were presented in the studied order with a blank replacing each target (e.g., “The emotional or _________ part.”).

**Procedure**

Participants were tested in small groups. They were told they would study two paragraphs on different topics, one sentence at a time, and that each paragraph would be followed by an unspecified memory test. They were also told that one word in each sentence (except the first two sentences) would include a word in red, and they should try to remember these words. The groups receiving generate sentences were told that the target word in red would (sometimes) be presented as a word fragment to be solved. An example generate sentence was provided.

In Block 1, the within group studied a paragraph containing 5 read sentences and 5 generate sentences in an alternating order that was counterbalanced across participants. The read group received 10 read sentences and the generate group received 10 generate sentences. Sentences appeared in turn, for 15 s each, on a large screen. Participants wrote each target down on a separate page of a response booklet. After the paragraph was presented participants did a math distractor task for 2 min. They then had 2 min to complete a fill-in-the-blank test sheet. Block 2 was then conducted identically using the other paragraph. Testing took about 30 min.

Half of each group received a study strategy questionnaire after Block 2. For the within group, the questionnaire began with a question assessing their awareness of the generation effect during Block 1: “What did you notice about how you did on the test of the *first* paragraph?”.
Each group then retrospectively self-reported their study strategy for each experienced target type in Block 1 and 2 in turn. Participants were also asked a yes/no question regarding whether they had changed study strategies across blocks.

Results

Following de Winstanley and Bjork (2004), a target was counted as recalled only if it was recalled in the correct sentence. A more lenient scoring method, where a target was counted as recalled if it was written in any blank produced higher recall but otherwise the same pattern (see also Bjork & Storm, 2011). Effects were significant at .05 level unless otherwise stated.

Recall in the within and between groups

Figure 1 displays the mean proportions of correct target recall. We first evaluated whether the within group replicated de Winstanley and Bjork (2004, Experiment 1). The within group’s recall was analyzed using a 2 (target type: read, generate) x 2 (test: 1, 2) repeated-measures ANOVA. Recall was higher for generate than read targets (.48 vs. .38; a generation effect), $F(1, 95) = 19.19, MSE = .04, \eta^2 = .17$, and recall increased from Test 1 to 2 (.37 vs. .49), $F(1, 95) = 24.89, MSE = .05, \eta^2 = .21$. These main effects were qualified by a significant interaction, $F(1, 95) = 7.72, MSE = .06, \eta^2 = .01$. The generation effect was significant on Test 1 (.48 vs. .29), $t(95) = 4.97, SE = .03, d = .51$, but not on Test 2 (.50 vs. .45), $t(95) = .79, SE = .03, p = .43, d = .08$. Recall increased significantly across tests for read targets (.29 vs. .48), $t(95) = 6.16, SE = .03, d = .62$, but not for generate targets (.45 vs. .50), $t(95) = 1.40, SE = .04, p = .16, d = .14$. This exact replication of de Winstanley and Bjork confirms that the elimination of the generation effect on Test 2 reflected increased read recall, not decreased generate recall.

Next, we compared recall of read targets in the within group (who experienced the generation manipulation) to the read group (who did not). A 2 (group: within, read) x 2 (test: 1,
2) mixed-factor ANOVA revealed higher read target recall in the read group than in the within group (.47 vs. .38), $F(1, 190) = 8.42, MSE = .09, \eta^2 = .81$ (perhaps due to some “lazy reading” of read sentences in the within group; see Begg & Snider, 1987), and a general increase in recall across tests (.35 vs. .50), $F(1, 190) = 40.61, MSE = .05, \eta^2 = .18$. Critically, the interaction was not significant: The increase in recall of read targets across tests was similar for the within and read groups, $F(1, 190) = 2.10, MSE = .05, p = .15, \eta^2 = .01$. Indeed, recall increased across tests in the read group who did not experience the generation manipulation (.41 vs. .52), $t(95) = 3.19, SE = .04, d = .38$. Clearly, then, an improvement for read targets on Test 2 is not uniquely due to experiencing a generation effect on Test 1.

Analogously, recall of generate targets was compared across the within and generate groups. Recall was similar for generate targets in the within and generate groups (.48 vs. .44), $F(1, 190) = 1.41, MSE = .08, p = .24 \eta^2 = .01$, and recall increased across tests (.43 vs. .49), $F(1, 190) = 7.90, MSE = .05 \eta^2 = .04$. Although the interaction was not significant, $F < 1$, generate target recall increased across tests in the generate group (.41 vs. .48), $t(95) = 3.07, SE = .02, d = .30$, but not in the within group (as reported above). Thus, the elimination of the generation effect on Test 2 in the within group reflected in part a hidden cost to generate targets. Experiencing a generation manipulation during Block 1 had a negative influence on generate targets in Block 2.

For completeness, the pure read and generate groups were also compared using a 2 (group: read, generate) x 2 (test: 1, 2) mixed-factor ANOVA. Recall increased across tests (.41 vs. .50), $F(1, 190) = 17.22, MSE = .05, \eta^2 = .08$ but was very similar in the read and generate groups (.46 vs. .45), $F < 1$, and there was no interaction with test, $F < 1$. Thus, a between-subject generation effect was absent (cf. de Winstanley & Bjork, 2004; Experiment 3), though this is
certainly not unprecedented (e.g., Begg & Snider, 1987; Begg, Vinski, Frankovich, & Holgate, 1991; de Winstanley & Bjork, 1997; de Winstanley, Bjork & Bjork, 1996).

**Self-reported study strategies**

Our post-experiment questionnaire explicitly asked participants whether they shifted study strategies across tests. We also asked them to describe the study strategy used for each target type they received on each test. These reports allowed us to: (1) identify whether participants reported shifting strategies across tests, (2) identify the strategies they used, (3) evaluate whether the within group was particularly likely to shift to a context strategy for read targets during Study 2 (Bjork & Storm, 2011), and (4) identify whether use of other strategies increased across blocks.

On the yes/no question, 54% of the within group and 81% of each of the read and generate groups reported shifting study strategy across blocks, suggesting improved recall across tests was due to more than a passive benefit of practice at taking a fill-in-the-blank test. Moreover, shifting strategies across blocks was clearly not unique to the within group, something that could not be detected in studies relying on within-subject strategy manipulations (de Winstanley & Bjork, 2004; Hertzog et al., 2008; 2009).

A coding scheme based on categorical groupings of participants’ reports revealed five study strategy categories: (1) linking targets with their sentences (*context*), (2) thinking about target meaning and/or mentally imagining targets (*elaboration*), (3) silently repeating targets (*repetition*), (4) focusing solely on writing targets down (*target only*), and (5) *no strategy*. Table 1 provides a breakdown of each group’s reported strategies. Participants tended to use the “shallower” strategies (target only, repetition, no strategy) over “deeper” strategies (context, elaboration). Most reported using different strategies for the two tests, though their reports did
not always align with their yes/no response. Given the small number of participants reporting each strategy we could not compare recall as a function of strategy type. Thus, the following evaluations of the strategy reports are purely descriptive.

Bjork and Storm (2011) suggested that the within group’s improvement for read targets across tests was due to increased sentence-target linking during Study 2. Consistent with this possibility, their participants were more likely to switch to a context strategy on Test 2 when Test 1 was a fill-in-the-blank test (which highlights the importance of the sentence frames) rather than free recall (which does not). Concordant with their finding, reported use of a context strategy for studying read sentences increased from Block 1 to 2 in our within group (.14 vs. .30). Reported use of a context strategy similarly increased across blocks for generate targets (.16 vs. .39), even though recall of generate targets did not increase. Additionally, use of a repetition strategy also increased across blocks, both for read targets (.27 vs. .39) and generate targets (.22 vs. .34). This parallel increase in the use of two different strategies makes it difficult to cleanly attribute the Test 2 improvement for read targets to increased use of a context strategy. In addition, context strategy use also increased slightly across blocks in the read group (.27 vs. .33), and quite dramatically in the generate group (.21 vs. .44), showing that paying more attention to the sentence frames during Study 2 was ubiquitous.

**Discussion**

The within group fully replicated de Winstanley and Bjork (2004), but their improvement in read target recall across tests was similar to a pure read group who never experienced the generation manipulation. A pure generate group also showed improved recall across tests. This pattern suggests that after Block 1 all groups made a similar strategy shift. Participants’ strategy reports suggest that all groups shifted their strategy during Study 2 toward forging sentence-
target links. The general increase in the use of a context strategy across blocks suggests participants may have learned how to better study the paragraphs for a fill-in-the-blank test (see also Bjork & Storm, 2011). Therefore, learning about the effects of generation during Test 1 may be less important than learning how to study for a particular type of memory test. However, use of a target repetition strategy also increased across blocks, making it hard to cleanly attribute changes in recall to a context strategy shift. Participants may have used different (or multiple) strategy shifts to achieve higher recall on Test 2.

The present results also provide an interesting challenge to the strategy knowledge updating framework (Dunlosky & Hertzog, 2000), in that the pure read and generate groups appear to have monitored their performance, updated their strategy knowledge, and utilized a new strategy for Block 2 despite receiving only one nominal study strategy. Our pure list group results appear to violate the framework’s monitoring assumption that:

“An individual must accurately monitor the differential effectiveness of the strategies during the task, either while studying or while being tested. According to this assumption, knowledge updating cannot occur if an individual is unaware that learning or performance differs for items studied with different strategies.”

(Dunlosky & Hertzog, 2000, p. 463).

Instead, our results suggest that a study strategy shift can be induced even in the absence of a formal comparison strategy.

Finally, relative to a pure generate group, Experiment 1 also revealed that the elimination of the generation effect on Test 2 in the within group was due in part to impaired generate target recall. Experiencing a generation manipulation in Block 1 may lead to a trade-off in Block 2 toward the read sentences, an effect that was not detected by de Winstanley and Bjork (2004).
Experiment 2

Experiment 2 tested de Winstanley and Bjork’s (2004) claim that participants learn about the memorial benefits of strategy effectiveness during the initial test. In addition, Hertzog et al. (2008) suggested that “monitoring strategy use during study apparently does not contribute greatly to knowledge about differential strategy effectiveness” (p. 443). More specific to the present paradigm, Bjork and Storm (2011) suggested “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 experience [to] develop more effective strategies for the encoding of future to-be-learned information” (p. 1122).

Although having a test experience can foster learning, it is likely that under some conditions participants will realize during Study 1 that they are encoding one type of target better than another. If so, their test experience would merely confirm their hypothesis. To directly compare these possibilities, a replication of the within group from Experiment 1 was compared to a new within group that did not receive Test 1. This no-Test-1 group had no opportunity to experience the relative benefits of generation over reading on a test. Therefore, if an initial test experience is essential for knowledge updating and strategy use shifting, then the generation effect should persist on Test 2 in the no-Test-1 group.

Method

Additional participants were randomly assigned to either a within group or a no-Test-1 within group (n = 48 each). The materials and procedure were the same as Experiment 1, except the no-Test-1 group did 2 min of a Sudoku distractor task after the math distractor task and before Block 2, rather than taking a 2 min Test 1.

Results
As shown in Figure 1, the within group pattern perfectly replicated Experiment 1 and de Winstanley and Bjork (2004). There was a generation effect (.43 vs. .35), $F(1, 47) = 8.22$, $MSE = .04$, $\eta^2 = .15$, recall increased from Test 1 to 2 (.35 vs. .44), $F(1, 47) = 7.15$, $MSE = .06$, $\eta^2 = .13$, and these main effects were qualified by a significant interaction, $F(1, 47) = 7.86$, $MSE = .04$, $\eta^2 = .14$. The interaction reflected a significant generation effect on Test 1 (.43 vs. .27), $t(47) = 4.06$, $SE = .04$, $d = .67$, but not on Test 2 (.44 vs. .44), $t < 1$. Recall increased across tests for read targets (.27 vs. .44), $t(47) = 4.05$, $SE = .04$, but not for generate targets (.43 vs. .44), $t < 1$.

Is experiencing the generation effect on an initial test necessary to spur participants to modify their encoding strategies in ways that lead to its subsequent elimination? If so, then a generation effect should persist on Test 2 in the no-Test-1 group. A 2 (target type: read, generate) x 2 (group: within, no-Test-1 within) mixed-factor ANOVA revealed no difference in recall for read and generate targets (.38 vs. .40), and no interaction with group, $F$s < 1. Thus, learners do not need to experience an effect of generation on an initial test to spawn strategy shifts. Interestingly, however, Test 2 recall was greater in the within group than the no-Test-1 group (.44 vs. .34), $F(1, 94) = 4.39$, $MSE = .11$, $\eta^2 = .05$. Therefore, taking an initial test clearly provides a benefit to recall above and beyond what an initial study experience provides.

**Discussion**

The generation effect was eliminated on Test 2 even when there was no Test 1, challenging de Winstanley and Bjork’s (2004) claim that experiencing the generation effect during Test 1 sows the seed of its elimination on Test 2 (cf. also Hertzog et al., 2008). Instead, participants may experience differential encoding of generate versus read targets during Study 1 and shift toward learning the worse target type during Study 2 (or may shift generally toward
greater encoding of sentence-target links; see Experiment 1). Experiment 3 employed metamemory measures to more closely examine this possibility.

In addition, Experiment 2 showed that taking Test 1, though not critical for eliminating the generation effect, led to greater recall on Test 2. Taking an initial test (cf. Roediger & Karpicke, 2006) could “boost” strategy shifting. For example, participants who experience a fill-in-the-blank test may be especially likely to increase their use of a context strategy in Block 2. Alternatively, this effect could reflect a general benefit of practice taking a given memory test.

**Experiment 3**

In Experiment 3, we measured participants’ metamemorial awareness of the effects of the generation manipulation at several stages of the procedure, a research direction recommended by Bjork et al. (2007) in a review of their research program using this paradigm. To this end, four metamemory measures (after Dunlosky & Hertzog, 2000; Hertzog et al., 2008, 2009) were collected in a new within group to test whether participants experience the effects of generation primarily during Test 1 (as per de Winstanley & Bjork, 2004; Hertzog et al., 2008), or whether they are sensitive to differences between their encoding of the two target types prior to Test 1 (as suggested by the no-Test-1 group in Experiment 2).

Two metamemory measures preceded Test 1. First, participants provided *judgments of learning (JOLs)* during Study 1 to gauge their monitoring of their learning. If participants are sensitive to the effects of the generation manipulation prior to Test 1, as our Experiment 2 suggests, then their JOLs for read and generate targets would be expected to differ at this point. Second, participants made quantitative recall *predictions* for each target type after Study 1 but before Test 1—what Hertzog et al. (2009) termed “poststudy differentiated predictions”. If
participants gain explicit knowledge about the effects of generation on their encoding during Study 1 then their predictions should be differentiated and hence should reveal this knowledge.

To evaluate memory performance monitoring during Test 1, confidence judgments (CJs) were collected for each recalled target. If participants become sensitive to the effects of generation only during Test 1 then these CJs should reveal this emergent awareness. Finally, to assess participants’ sensitivity to strategy efficacy they made quantitative recall postdictions for each target type after Test 1—what Dunlosky and Hertzog (2000) termed “global-differentiated postdictions”. These postdictions should indicate whether participants experienced differential recall for generate and read targets on Test 1.

The same four metamemory measures were also collected in Block 2. Given the results of Experiments 1 and 2, it was expected that the generation effect would be eliminated on Test 2; thus metamemory differences for read and generate targets were not expected or examined. However, we did investigate whether shifts in recall for each target type across blocks coincided with shifts in metamemory reports, given that metacognitive studies (e.g., Dunlosky & Hertzog, 2000; Koriat & Goldsmith, 1996; Nelson & Narens, 1990) typically examine the relationships between monitoring (assessed via metamemory measures) and control (assessed via memory performance).

Finally, as detailed below, recall and metamemory measures were analyzed at the level of subgroups who differed in their Test 1 recall pattern. These subgroup analyses may provide a potentially important new window into the learning processes across study-test experiences.

**Method**

An additional 48 University of Calgary undergraduates were tested. A collection of four metamemory measures was added to the usual within group procedure. During Study 1,
Learning Across Study-Test Experiences 19

participants made a JOL after each sentence was studied. They wrote down a number between 0 and 100 to indicate their confidence that they would remember that target at test (higher number = greater confidence). After Study 1, but before Test 1, participants predicted the number of generate targets out of 5, and the number of read targets of out 5, they thought they would recall. During Test 1, participants wrote a CJ next to each target they recalled by assigning a number between 0 and 100 (higher number = greater confidence); CJs were analyzed only for correctly recalled targets though analyzing CJs for all responses did not affect the pattern of results. Finally, after Test 1 participants postdicted the number of read targets out of 5, and the number of generate targets out of 5, they thought they recalled. This procedure was repeated for Block 2.

Results and Discussion

After analyzing recall in the within group we segue to an important set of cross-experiment analyses to establish that our within groups consisted of three distinct subgroups differing in their Test 1 recall pattern. We then analyze our metamemory measures at this subgroup level to determine whether and when each subgroup became aware of their recall pattern during Block 1. Finally, where subgroups showed metamemorial differences for read and generate targets on Test 1, we examined whether and how their metamemory responses shifted across blocks.

Recall in the within group

As shown in Figure 1, recall for read and generate targets was equivalent overall (.35 vs. .36), $F < 1$, and recall increased from Test 1 to Test 2 (.29 vs. .43), $F(1, 47) = 22.07, MSE = .04, \eta^2 = .32$. The usual target type by test interaction was marginal, $F(1, 47) = 3.00, MSE = .05, p = .09, \eta^2 = .06$, but the pattern was the same as in Experiments 1 and 2: a generation effect was present on Test 1 though here it was only marginal (.25 vs. .32), $t(47) = 1.76, SE = .04, p = .08, d$
= .25, but not on Test 2 (.45 vs. .40), \( t(47) = 1.08, SE = .05, p = .29, d = .16 \). Recall of read targets nearly doubled across tests (.25 vs. .45), \( t(47) = 4.72, SE = .04, d = .69 \), whereas the increase for generate targets did not reach significance (.32 vs. .40), \( t(47) = 1.67, SE = .05, p = .10, d = .24 \). Although collecting metamemory measures during study and test somewhat diluted the overall recall pattern, critically, the generation effect was once again absent on Test 2.

**Cross-experiment subgroup analyses**

We reasoned that different Test 1 recall patterns would likely be associated with different metamemory patterns. For example, participants who show a generation effect on Test 1 should typically report higher metamemory responses for generate targets than read targets, whereas this concordance might not be true of participants who do not show a generation effect on Test 1. To examine this possibility, we therefore split the within groups from Experiments 1 to 3 into three subgroups based on the direction of their objective Test 1 recall pattern: the \( R < G \) subgroup \((n = 54, 28, 21 = 103 \text{ total})\) recalled numerically more generate targets than read targets on Test 1 (i.e., generation effect), the \( R > G \) subgroup \((n = 20, 7, 14 = 41 \text{ total})\) recalled numerically more read targets than generate targets on Test 1 (i.e., negative generation effect), and the \( R = G \) subgroup \((n = 22, 13, 13 = 48 \text{ total})\) showed equal recall of read and generate targets on Test 1.

Although the within-subject generation effect is typically robust (see Bertsch et al., 2007), sizeable \( R > G \) and \( R = G \) subgroups were not unexpected. To succeed on a fill-in-the-blank test, participants need to recall the target given its sentence as a cue, thus performance on this test is likely to benefit from relational processing (e.g., Einstein & Hunt, 1980)—in this case sentence-target linking. Some participants likely used the sentences to help them solve the fragments (perhaps the \( R < G \) subgroup in particular), whereas others may have focused largely on item-specific processing of the fragments themselves (perhaps the \( R > G \) subgroup in particular).
Participants who ignored the sentence frames for generate sentences would have difficulty succeeding on a fill-in-the-blank test for those targets. In general, participants who performed more relational processing for read (vs. generate) sentences, and/or more item-specific processing for generate (vs. read) sentences, might not show a generation effect on this test.

Test 2 recall across subgroups was compared using a 3 (subgroup: R>G, R<G, R=G) x 2 (target type: read, generate) mixed-factor ANOVA (see Figure 3). The subgroup pattern was very similar in each experiment (see Burnett, 2013, for these analyses). Experiment did not interact with these factors in preliminary analyses so it was not included as a factor here. Overall Test 2 recall was not similar across the R<G, R>G, and R=G subgroups (.46 vs. .49 vs. .46), $F(2, 189) = 1.79$, $MSE = .089$, $p = .17$, $\eta^2 = .01$. Most importantly, there was no hint of a generation effect on Test 2 in the pooled within group (.46 vs. .46), $F < 1$, nor was there an interaction of subgroup and target type, $F < 1$. The generation effect was not significant on Test 2 in the R<G subgroup (.45 vs. .47), $t < 1$, the R>G subgroup (.53 vs. .45), $t(40) = 1.67$, $SE = .05$, $p = .10$, $d = .40$, or the R=G subgroup (.43 vs. .46), $t < 1$. Thus, the pattern of recall on Test 1 did not reliably influence the pattern of recall on Test 2, contrary to de Winstanley and Bjork’s (2004) claim that the elimination of the generation effect on Test 1 reflects experiencing this effect on Test 1.

We next examined how recall in each subgroup shifted across tests using 2 (target type: read, generate) x 2 (test: 1, 2) repeated-measures ANOVAs. In the R<G subgroup, recall increased across tests (.32 vs. .46), $F(1, 102) = 20.96$, $MSE = .05$, $\eta^2 = .17$, but this was qualified by a significant interaction with target type, $F(1, 102) = 114.87$, $MSE = .03$, $\eta^2 = .53$. Recall of read targets more than doubled across tests (.18 vs. .47), $t(102) = 10.76$, $SE = .03$, $d = 1.26$, whereas recall of generate targets decreased (.54 vs. .45), $t(102) = 3.17$, $SE = .03$, $d = .39$. Thus, participants who experienced the generation effect on Test 1 eliminated it on Test 2 by becoming
“better readers” as de Winstanley and Bjork (2004) claimed. However, this benefit also came with a cost to generate targets that was not detected in previous studies. Therefore, this subgroup also became “worse generators” so to speak.

In the R>G subgroup, recall again increased across tests (.33 vs. .49), $F(1, 40) = 21.47$, $MSE = .05$, $\eta^2 = .35$, and this was again qualified by a significant interaction with target type, $F(1, 40) = 42.38$, $MSE = .03$, $\eta^2 = .51$. Here, recall more than doubled across tests for generate targets (.20 vs. .53), $t(40) = 7.11$, $SE = .05$, $d = 1.43$, but did not change across tests for read targets (.46 vs. .45), $t < 1$. Thus, participants who experienced a negative generation on Test 1 eliminated it on Test 2 by becoming “better generators” but they did not correspondingly become “worse readers”—likely because reading words is such a well-practiced skill. Importantly, both of these subgroups appear to have shifted their encoding toward their more poorly recalled type of target on Test 1. Experiencing a generation effect is not the only Test 1 outcome that spurs recall shifts across tests—the same can be true of experiencing a negative generation effect!

Recall in the R=G subgroup also increased across tests (.32 vs. .45), $F(1, 47) = 11.05$, $MSE = .07$, $\eta^2 = .19$, but here the interaction with target type was not significant, $F < 1$; instead, recall increased across tests for both read targets (.32 vs. .46), $t(47) = 3.52$, $SE = .04$, $d = .60$, and generate targets (.32 vs. .43), $t(47) = 2.35$, $SE = .05$, $d = .48$. Participants who did not experience differential Test 1 recall improved for both types of targets on Test 2, perhaps because they had no incentive to shift their focus to one target type over another during Study 2. Thus, either this subgroup became “better readers” and “better generators”, or learned to emphasize a context strategy in anticipation of another fill-in-the-blank test.

**Metamemory measures**
The previous section confirmed the existence of three within-group subgroups, each of which made a unique shift in recall for the two target types across tests. Given the occurrence of three different recall patterns on Test 1, we analyzed our four metamemory measures at the subgroup level (see Figure 3). We first examined the metamemory measures collected during Block 1 to determine whether/when each subgroup became aware of their eventual Test 1 recall pattern. We then examined how each subgroup’s metamemory shifted across blocks. Because each subgroup showed similar recall for read and generate targets on Test 2, we did not separately analyze the metamemory measures for Block 2.

**Metamemory during Block 1.** Study 1 JOLs (Figure 3, Panel A) were similar for read and generate targets in each subgroup: R<G subgroup (76 vs. 76), t < 1, R>G subgroup (71 vs. 68), t < 1, and R=G subgroup (70 vs. 66), t(12) = 1.81, SE = 2.23, p = .10, d = .19. Predicted recall of read and generate targets collected after Study 1 (Figure 3, Panel B) was also similar for each subgroup: R<G subgroup (70% vs. 74%), t(20) = 1.28, SE = .15, p = .21, d = .23, R>G subgroup (63% vs. 60%), t<1, and R=G subgroup (68% vs. 63%), t(12) = 1.15, SE = .20, p = .27, d = .28. Assuming the validity of these measures, the R<G and R>G subgroups’ reports during and after Study 1 showed little sensitivity to their eventual Test 1 recall pattern. This lack of awareness is surprising given the generation effect on Test 2 was eliminated in Experiment 2 when there was no Test 1 (which suggested learning occurs during or after Study 1).²

In contrast to the first two metamemory measures, CJs collected during Test 1 (Figure 3, Panel C) were much lower for read than generate targets in the R<G subgroup (32 vs. 72), t(20) = 4.41, SE = 9.00, d = 1.13, and conversely were much higher for read than generate targets in the R>G subgroup (84 vs. 51), t(13) = 2.60, SE = 12.70, d = .94. Thus, both subgroups reported differences in confidence during Test 1 concordant with their actual recall pattern (see Figure 2).
Importantly, however, these CJ differences do not necessarily imply an awareness of having shown either a generation effect (R<G subgroup) or a negative generation effect (R>G subgroup). Participants did not identify whether each recalled, rated target was generated or read at study. Therefore, although participants had higher confidence for some recalled targets over others, there is no evidence that they knew whether those targets had been generated or read at study. CJ's in the R=G subgroup were similar for read and generate targets (76 vs. 80), $t < 1$.

Consistent with the possibility that each subgroup’s CJ pattern did not reflect awareness of the effect of generation during Test 1, postdicted recall after Test 1 (Figure 3, Panel D) did not differ significantly for read and generate targets in the R<G subgroup (37% vs. 39%), $t < 1$, R>G subgroup (40% vs. 31%), $t(13) = 1.71, SE = .25, p = .11, d = .43$, or R=G subgroup (40% vs. 34%), $t(12) = 1.30, SE = .24, p = .22, d = .31$.

In summary, participants did not report metamemorial differences between read and generate targets during or after Study 1, or even after Test 1. Moreover, the sole metamemory differences we observed (CJs during Test 1) do not necessarily indicate an awareness of the relative effects of generation versus reading—they may merely indicate a difference in confidence, from participants’ perspectives, in the recall of some targets relative to others. Overall, participants’ lack of metamemorial sensitivity suggests that Block 1, rather than teaching participants about the effects of generation, might simply teach them how to better study to succeed on a fill-in-the-blank test. If so, then an important lesson from our experiments is that it may be more profitable for students to attend to the demands of a memory test rather than to their relative success at encoding and retrieving different types of items.

**Metamemory shifts across blocks.** Because the only differences in metamemory for read and generate targets in Block 1 occurred for CJs, we limited our analyses of recall shifts
across blocks to this measure. Specifically, we tested whether the shift in the R<G and R>G subgroups’ CJs across blocks (Figure 3) mirrored their mutual shifts to a null generation effect (Figure 2), using 2 (target type: read, generate) x 2 (test: 1, 2) repeated-measure ANOVAs.

In the R<G subgroup, CJs increased from Test 1 to Test 2 (52 vs. 73), $F(1, 20) = 9.00$, $MSE = 1,002, \eta^2 = .31$, and the key interaction with target type was significant, $F(1, 20) = 11.04$, $MSE = 682, \eta^2 = .36$. CJs more than doubled across tests for read targets (32 vs. 72), $t(20) = 4.00$, $SE = 10.00, d = 1.16$, but not for generate targets (72 vs. 74), $t < 1$. Aligned with the similar recall of read and generate targets on Test 2, there was no difference in CJs on Test 2 (72 vs. 74), $t < 1$. Thus, both CJs and actual recall in the R<G subgroup increased across tests for read targets but not for generate targets.

In the R>G subgroup, CJs were similar across Test 1 and 2 (68 vs. 80), $F(1, 13) = 2.34$, $MSE = 1,009, p = .16, \eta^2 = .15$, and the interaction was marginally significant, $F(1, 13) = 3.82$, $MSE = 859, p = .07, \eta^2 = .23$. The increase in CJs across tests was marginally significant for generate targets (51 vs. 79), $t(13) = 1.96, SE = 14.29, p = .07, d = .70$, but not for read targets (84 vs. 81), $t < 1$. In line with Test 2 recall, CJs for read and generate targets were similar on Test 2 (81 vs. 79), $t < 1$. Although the R<G subgroup was small ($n = 14$), there was some evidence of a parallel increase in both recall and CJs for generate targets but not for read targets across tests.

In sum, differences in CJs for read and generate targets on Test 1 do not necessarily indicate explicit awareness of a generation or negative generation effect. On the other hand, these CJs mapped onto recall shifts across blocks in the manner that awareness would predict. Participants who fared worse on read targets on Test 1 (R<G subgroup) showed an selective increase in CJs for read targets on Test 2, and those who fared worse on generate targets on Test 1 (R>G subgroup) showed a marginal selective increase in CJs for generate targets on Test 2.
General Discussion

We examined how and when participants learn about the effects of a generation study strategy across study-test experiences. We obtained three successful replications of de Winstanley and Bjork’s (2004; see also Bjork & Storm, 2011) elimination of the within-subject generation effect across fill-in-the-blank tests. Each of our experiments also added an important new finding that tempers some of the conclusions reached in prior studies regarding strategy knowledge updating (Bjork & Storm, 2011; de Winstanley & Bjork, 2004; Hertzog et al., 2008; 2009). In Experiment 1, a pure read group improved as much as a within group across tests even though the former group did not experience the generation effect. Thus, participants can learn to modify their study strategies without experiencing two strategies that vary in effectiveness. In Experiment 2, a within group who did not receive Test 1 nonetheless showed the same elimination of the generation effect on Test 2. Thus, knowledge updating can occur at study not just during a memory test. Finally, in Experiment 3 knowledge updating across study-test experiences appears to have occurred without metamemorial awareness. Below we unpack these findings and how they add to our understanding of encoding, recall, and metamemory.

Learning across study-test experiences

de Winstanley and Bjork (2004) and Bjork and Storm (2011) claimed that learning about the generation effect during Test 1 led to increased use of a context strategy that worked to eliminate the effect on Test 2. Both studies reported evidence consistent with these claims, but our findings suggest they were too restrictive in four ways.

1. Pure-list groups can also improve across tests. In Experiment 1, the increase in recall of read targets was not significantly greater in the within group (who experienced the generation effect) than in the pure read group (who did not). Recall of generate targets also
increased across tests in the pure generate group. Thus, factors other than monitoring and updating knowledge about the effects of an encoding manipulation operate in this paradigm. That all groups in Experiment 1 showed some improvement on Test 2 suggests that during Block 1, participants may have learned about the importance of encoding the sentence contexts for succeeding on a fill-in-the-blank test. Consistent with this possibility, reported use of a context strategy during Study 2 increased for all groups. Across study-test experiences, participants may learn how best to study for a given memory test rather than learning about the relative benefits of a given study strategy per se.

2. Study strategies typically shift across blocks. Participants who experienced the generation effect on Test 1 (R\textless{}G subgroup) improved their recall of read targets on Test 2, whereas those who experienced a negative generation effect on Test 1 (R\textgreater{}G subgroup) improved their recall of generate targets on Test 2. Critically, neither subgroup showed a generation effect on Test 2. Thus, participants appear to have shifted their encoding efforts to whichever target type suffered during Block 1, though it remains debatable to what extent they were aware of this shift (Experiment 3). Curiously, participants who did not experience differential recall on Test 1 (R\textequal{}G subgroup) improved their recall of both types of targets on Test 2 (and they too did not show a generation effect on Test 2).

Bjork and Storm (2011, Experiment 4) collected metamemory reports that suggested the within group’s improved recall of read targets on Test 2 was due to increased use of a context strategy during Block 2. In addition, they showed that recall of context words on Test 2 was greater when Test 1 was fill-in-the-blank rather than free recall. However, the authors did not establish a direct link between use of a context strategy and increased target recall. Thus, it
remains possible that a context strategy was not the source of the recall shift, or that increased use of a context strategy increased sentence context recall but not target recall.

Our within group in Experiment 1 was also more likely to report using a context strategy for read sentences during Study 2 (vs. Study 1). However, this increased use of a context strategy was also true of generate targets, which would not explain why generate recall did not also increase across tests thus preserving the generation effect on Test 2. We also found that use of a repetition strategy increased across blocks. Because the within group increased their use of more than one strategy across blocks the elimination of the generation effect cannot be cleanly attributed to the increased use of a context strategy. In addition, because the between groups in Experiment 1 also increased their use of a context strategy across blocks, this shift may reflect a general reaction to completing a fill-in-the-blank test (cf. de Winstanley & Bjork, 2004).

Interestingly, the majority of each group in Experiment 1 reported changing study strategies across blocks, suggesting learners routinely update their study strategies after task experience (Bjork, 2011; Delaney, Verkoeijen, & Spirgel, 2010; Dunlosky & Hertzog, 2000; Schmidt & Bjork, 1992). Unfortunately, there were too few participants to enable us to test whether recall and recall shifts differed as a function of strategy. Future studies could assign participants a particular study strategy for each block to assess these questions. Another approach is to have participants rate how much they used each of several strategies, and to then use regression to evaluate the relationship between strategy use and recall (see Hertzog et al., 2008).

3. Shifting study strategies can also yield costs. de Winstanley and Bjork (2004) found similar recall of generate targets across tests. In contrast, we found that within group participants who experienced the generation effect on Test 1 (i.e., the R<G subgroup) showed a decrease in recall of generate targets across tests. This implies a trade-off strategy in which increased
encoding of read items led to impaired encoding of generate items. Consistent with this pattern is Begg and Snider’s (1987) finding that read targets are sometimes remembered more poorly than generate targets due to the processing demands of generation. That is, the within generation effect can sometimes reflect “lazy reading”. Our results suggest that in some cases, participants may become “lazy generators” in an effort to improve their recall of read targets. This potential cost effect to learning about the effectiveness of a study strategy warrants further study.

4. Learning from a study strategy can occur at study, not just at test. Learning resulting from experience with a study strategy has largely been posited to arise during a memory test (Bjork & Storm, 2011; de Winstanley & Bjork, 2004; Hertzog et al., 2008; 2009). In the current paradigm, de Winstanley and Bjork (2004; Bjork & Storm, 2011) emphasized that Test 1 was the origin of the elimination of the generation effect on Test 2. Contradicting the generality of this claim, the within group in Experiment 2 who did not receive a Test 1 also did not show a generation effect on Test 2. Participants can sometimes learn something about the effectiveness of a study strategy during study, although as discussed below our metamemory measures in Experiment 3 did not reveal much evidence of explicit learning. The monitoring assumption in the strategy knowledge updating framework should thus be updated to recognize a role for monitoring strategies that occur during study (Dunlosky & Hertzog, 2000).

It is important to emphasize that acknowledging a role for learning about study strategy efficacy during study does not rule out a role for additional learning at test. For example, in Experiment 2, Test 2 recall was higher when Test 1 was presented (vs. not presented). This difference could simply reflect task practice, but it may well also reflect additional knowledge updating gained from taking an initial test. In our paradigm, participants taking Test 1 may learn the value of attending to the sentence frames for succeeding on a fill-in-the-blank test. In
addition, in Experiment 3, CJs collected during Test 1 mapped neatly onto each subgroup’s recall shift, suggesting that some form of learning (whether explicit or implicit) occurred during Test 1.

**Implications for accounts of the generation effect**

Although our focus was on how encoding strategies, recall, and metamemory shift cross study-test experiences, our results will also need to be accommodated by accounts of the generation effect—none of which predict its elimination on Test 2. For example, the transfer-appropriate multifactor account (de Winstanley et al., 1996)—itself an updated version of the multifactor account (McDaniel et al., 1988)—predicts a generation effect on both tests because Test 2 should be just as sensitive as Test 1 to information that was strengthened by generation at study. On the procedural account (Crutcher & Healy, 1989; McNamara & Healy, 1995), when information is generated (versus read), people taking a test are more likely recapitulate or simulate the processing operations they had engaged in at during study. Both accounts incorrectly predict a generation effect on both tests in the present paradigm. Similarly, the distinctiveness account (Hunt, 2003; McDaniel & Geraci, 2006) suggests that the distinctively encoded generate targets should be better remembered than read targets, hence it also seems to predict a generation effect on both tests. To accommodate the elimination of the generation effect on Test 2, this account might, for example, posit that read targets during Study 2 were encoded as distinctively as generate targets, perhaps because a distinctive context strategy was used to encode both types of targets.

**Metamemory shifts (or lack thereof) across study-test experiences**

Surprisingly, Experiment 3 suggested that shifts in memory performance in this paradigm were not mirrored by shifts in metamemorial awareness. Based on the strategy knowledge updating framework (Dunlosky & Hertzog, 2000), knowledge updating should not occur if the
learner is unaware of a performance difference for items studied using different strategies (i.e., the monitoring assumption). We cannot rule out the possibility that our metamemory measures were not sensitive enough to detect differences in metamemory for read and generate items. Having only 5 trials per target type may have been insufficient, especially if participants need several trials to “tune in” to their metacognitive states. In addition, the less robust generation effect on Test 1 in Experiment 3 may have reduced participants’ metamemory for its effects (ironically due to our inclusion of metamemory measures!). Studies using JOLs (e.g., Dunlosky & Metcalfe, 2009) and predictions/postdictions (e.g., Hertzog et al., 2008) typically employ far more trials per condition (15-30) than our paragraph materials allowed. Indeed, in a different paradigm employing more trials Mazzoni and Nelson (1995, Experiment 2) found higher JOLs for generate targets than read targets at study, and JOLs were shown to mirror recall (see also Castel, McCabe, & Roediger, 2007).

However, other evidence suggests it may not be so surprising that our R<G and R>G subgroups in Experiment 3 did not provide different JOLs or predictions for generate and read targets during and after Study 1, respectively. For example, according to Koriat’s (1997) cue-utilization approach, JOLs can be based on intrinsic or extrinsic cues. Intrinsic cues include the characteristics of the target itself (e.g., meaning) and the perceived ease of remembering it on a subsequent test. Extrinsic cues include the encoding operations used to encode a target, such as a generation manipulation. Koriat found that people consistently ignored extrinsic factors when making JOLs. In addition, Froger, Sacher, et al. (2011) found that participants made lower predictions for recall of read (vs. generate) targets after a 5-item study phase using paired associates (though across separate testing sessions where the read session always occurred first). If our participants intuited that fragments would yield worse recall than intact words, this
intuition would need to be overcome to yield higher predicted recall for generate targets. Thus, it remains possible that the null effects in our prediction/postdiction metamemory measures could reflect a failure of participants’ experiences to overcome their initial intuitions.

Thus, it remains possible that our participants simply did not experience metamemorial differences for read and generate targets either at study or at test. If this inference proves valid, then the unique shifts in recall made by each subgroup may reflect a relatively tacit or implicit process rather than an explicit awareness. There is evidence that implicit strategy updating can precede explicit strategy knowledge updating, at least in children (e.g., Crowley, Shrager, & Siegler, 1997; Siegler & Stern, 1998). Moreover, as noted above, the CJ differences in Experiment 3 for read and generate targets, though provocative, do not demonstrate that participants were aware of their Test 1 recall pattern. To provide such a demonstration, future research could profitably ask participants to make a source-memory judgment for each recalled target along with a CJ, to test the link between recall and metamemory.

Where tacit or implicit changes in strategy knowledge updating can drive recall shifts across tests there may not be much educational benefit to encouraging learners to become aware of their relative recall for different target types on a test. In addition, as we have argued, learning how to meet the demands of a memory test may be at least as important as learning about the advantages of a given encoding strategy.

Conclusion

We provided a detailed examination of how learning about the benefits of one well-established study strategy—generation—affects encoding, recall, and metamemory and subsequent encoding, recall, and metamemory. Research has moved beyond measuring the immediate effectiveness of a given study strategy, toward an evaluation of when and how people
modify their learning strategies across study-test experiences. Such research has value for memory theory development as well as an applied value for learners and educators. We found that individuals can adaptively modify their learning across study-test experiences, perhaps without much awareness of doing so (Experiment 3). Nonetheless, as learners come to discover the benefits of a given study strategy, they create a scaffold for developing improved study strategies for themselves, and ultimately for other learners as well.
Footnotes

1. After Test 1, de Winstanley and Bjork (2004; Experiment 1B) asked their within group (N = 31) “What did you notice about your performance on the previous memory test?” Their aware subgroup (n = 17) reported awareness of the generation effect on Test 1, and 13 of the 17 (76%) did not have numerically higher recall for generate (vs. read) targets on Test 2. Their unaware subgroup (n = 14) did not report awareness of the generation effect on Test 1, and 8 of the 14 (57%) did not have numerically higher recall for generate (vs. read) targets on Test 2. Given the small sample size, this difference (76% vs. 57%) is not very compelling evidence for a role of awareness of the generation effect after Test 1 on its elimination on Test 2. We posed a similar question (e.g., “What did you notice about how you did on the test on the first paragraph?”), but only 1 of 48 participants reported awareness of the generation effect during Test 1. Our posing the question after Test 2 may have caused this low rate, due to forgetting or reporting errors.

2. Although it was not our focus, it is worth noting that both measures reveal a striking overconfidence relative to actual recall during Block 1 that was attenuated during Block 2.

3. Although the R<G and R>G subgroups were created on an objective basis (Test 1 recall), each subgroup each improved across tests for the target type they did worse on. We acknowledge that these shifts could reflect regression towards the mean rather than shifts in study strategy. However, a regression-toward-the-mean explanation would not predict our finding that CJs mapped onto each subgroup’s recall pattern. We also acknowledge that our subgroups could differ with respect to variables other than Test 1 recall. For example, the R<G (vs. R>G) subgroup may have had previous knowledge of the generation effect, or may be more test-experienced or engaged test-takers. The bases of the Test 1 subgroup patterns remain an interesting and open question.
Author Note

The Natural Sciences and Engineering Research Council of Canada supported this research through a Discovery Grant to GEB. This research formed part of ANB’s dissertation. Send correspondence to Glen Bodner, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary AB T2N 1N4, Canada (e-mail: bodner@ucalgary.ca).
References


Table 1

*Experiment 1: Proportion of each group reporting each study strategy by target type and test.*

<table>
<thead>
<tr>
<th>Group/Strategy</th>
<th>Read targets</th>
<th>Generate targets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Test 1</td>
<td>Test 2</td>
</tr>
<tr>
<td>Within group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>.14</td>
<td>.30</td>
</tr>
<tr>
<td>Elaboration</td>
<td>.07</td>
<td>.11</td>
</tr>
<tr>
<td>Repetition</td>
<td>.27</td>
<td>.39</td>
</tr>
<tr>
<td>Target only</td>
<td>.27</td>
<td>.11</td>
</tr>
<tr>
<td>No strategy</td>
<td>.25</td>
<td>.09</td>
</tr>
<tr>
<td>Between groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>.27</td>
<td>.33</td>
</tr>
<tr>
<td>Elaboration</td>
<td>.15</td>
<td>.17</td>
</tr>
<tr>
<td>Repetition</td>
<td>.17</td>
<td>.23</td>
</tr>
<tr>
<td>Target only</td>
<td>.35</td>
<td>.27</td>
</tr>
<tr>
<td>No strategy</td>
<td>.06</td>
<td>.02</td>
</tr>
</tbody>
</table>
Figure 1. Experiments 1-3: Mean proportion of targets recalled by target type, test, and group.

Error bars show standard errors.
Figure 2. Within group subgroups pooled across Experiments 1-3: Mean proportion of targets recalled by target type and test. Error bars show standard errors. The subgroup means for Experiment 3 are provided above each bar.
Figure 3. Experiment 3: Metamemory measure means by target type and test for each subgroup.

Error bars show standard errors.