Journal of Experimental Psychology:
Learning, Memory, and Cognition
2017, Vol. 43, No. 12, 1934–1947

0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000427

The Testing Effect Under Divided Attention

Zachary L. Buchin and Neil W. Mulligan
University of North Carolina at Chapel Hill

Memory retrieval often enhances later memory compared with restudying (i.e., the testing effect), indicating that retrieval does not simply reveal but also modifies memory representations. Dividing attention (DA) during encoding greatly disrupts later memory performance while DA during retrieval typically has modest effects—but what of the memory-modifying effects of retrieval? If these effects are similar to study-based encoding, they should be greatly disrupted by DA, a possibility consistent with elaborative and effortful accounts of the testing effect. Alternatively, the mnemonic consequences of retrieval may be largely resilient to distraction, like retrieval itself. In 3 experiments, participants studied word pairs (Phase 1) then engaged in restudy of some pairs and retrieval of others (Phase 2), followed by a final cued-recall test (Phase 3). Phase 2 restudy and retrieval occurred under full attention (FA) or DA. The experiments were designed to induce either material-specific (Experiments 1 and 2) or material-general (Experiment 3) interference, as well as to produce comparable secondary task performance between the restudy and retrieval groups (Experiments 2 and 3). Consistent with prior research, retrieval improved final recall (i.e., the testing effect) whereas DA disrupted final recall. Critically, the 2 factors interacted such that the negative effect of DA on final recall was substantial in the restudy condition but quite modest in the retrieval condition—resulting in a larger testing effect in the DA than FA condition. The encoding effects of retrieval seem resilient to distraction which has implications for theories of the testing effect.

Keywords: testing effect, divided attention, episodic memory

Researchers and educators have long studied memory to uncover effective methods by which to learn material. One useful technique is revealed by the testing effect, in which retrieval improves subsequent memory more than merely restudying the same material (or not reexperiencing it at all; Carpenter, Pashler, Wixted, & Vul, 2008; Carrier & Pashler, 1992; Chan & McDermott, 2007; Cull, 2000; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Zaromb & Roediger, 2010). Gates (1917) was one of the first researchers to find that recitation (i.e., repeated retrieval) improves memory, and numerous recent studies agree, finding that the testing effect arises in many situations, including lab studies, skill learning, multimedia stimuli tests, and classroom settings (Johnson & Mayer, 2009; Kromann, Jensen, & Ringsted, 2009; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Butler, 2011; Roediger & Karpicke, 2006).

The typical paradigm for exploring the testing effect consists of three phases, beginning with a study phase in which participants are presented with material (e.g., word pairs, educational text) to learn for a later test. In the second phase, participants either restudy the material or are given a test on the material. In the third phase, participants are given a final memory test. Typically, material tested in Phase 2 is better remembered than material that is restudied—the testing effect (e.g., Roediger & Butler, 2011).

This article was published Online First May 15, 2017.

Zachary L. Buchin and Neil W. Mulligan, Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill.

Correspondence concerning this article should be addressed to Zachary L. Buchin, Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, 235 East Cameron Avenue, Chapel Hill, NC 27599-3270. E-mail: buchinzl@live.unc.edu

The memory enhancement revealed by the testing effect shows that retrieval does more than simply reveal the contents of memory; it also modifies memory representations (Bjork & Bjork, 1992; Mulligan & Picklesimer, 2016; Roediger & Butler, 2011; Roediger & Karpicke, 2006). This mnemonic benefit of retrieval shows that retrieval has encoding (or perhaps, reencoding) effects. Consequently, it is important to determine how the encoding effects of retrieval operate especially compared to processes more typically labeled *encoding* (i.e., those that operate during periods of study or restudy).

In this light, the effects of attention on memory provide an interesting arena in which to compare the memory-modifying effects of retrieval and encoding. Dividing attention during encoding usually produces substantial negative effects on memory, whereas the effects of dividing attention during retrieval are typically quite modest (Anderson, Craik, & Naveh-Benjamin, 1998; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; LeCompte, Neely, & Wilson, 1997; Murdock, 1965; see Mulligan, 2008, for review). Although retrieval success is little impacted by divided attention, far less is known about the effects of divided attention on the subsequent mnemonic effects of retrieval. This domain is especially interesting because some theoretical perspectives imply that the mnemonic effects of retrieval should be quite sensitive to divided attention whereas other perspectives suggest less sensitivity. We begin with a brief review of relevant empirical results and then describe these contrasting perspectives.

To assess the effects of attention on memory, researchers often compare a divided attention (DA) condition, in which participants perform a memory task while also carrying out a secondary distractor task, and a full attention (FA) condition, in which only the memory task is performed (e.g., Craik et al., 1996; Lozito & Mulligan, 2010). When attention is divided during encoding, it is commonly found that later memory accuracy is greatly reduced (e.g., Anderson et al., 1998). However, dividing attention during the memory test often produces little or no decrement in accuracy (e.g., Craik et al., 1996). Even when DA significantly reduces test performance (e.g., Fernandes & Moscovitch, 2000; Hicks & Marsh, 2000; Mulligan & Lozito, 2006), the negative effects are smaller than DA effects during encoding (e.g., Anderson et al., 1998; Craik et al., 1996; Naveh-Benjamin, Craik, Perretta, & Tonev, 2000). Finally, although engaging in a secondary task during retrieval produces little effect on memory accuracy, performance on the secondary task itself is impaired by retrieval and to a greater degree than when the secondary task is paired with encoding (e.g., Anderson et al., 1998; Craik et al., 1996; Naveh-Benjamin et al., 2000).

Because of the differing effects of DA on encoding and retrieval, attention offers an interesting domain in which to examine the mnemonic effects of retrieval. If the encoding effects of retrieval are similar to the processes usually labeled *encoding*, then they should be quite sensitive to manipulations of attention. That is, dividing attention during Phase 2 retrieval might reduce final recall to a similar or even greater degree than the reduction expected to occur when attention is divided during restudy. This latter possibility is consistent with accounts of the testing effect which propose that the testing advantage is due to greater elaboration or effortful processing. For example, one account of the testing effect argues that retrieval enhances elaborative processing (Carpenter, 2009, 2011). According to this view, in attempting to retrieve a target, participants elaborate on the memory trace by activating semantic associates of the cue and to-be-retrieved target. These elaborations serve as additional retrieval routes for accessing the target information on still later memory tests (Carpenter & Yeung, 2017; Rawson, Vaughn, & Carpenter, 2015). Pyc and Rawson (2010) propose a similar, mediator-effectiveness hypothesis (see Rowland, 2014, for discussion). Research on attention and memory shows that elaborative processing is effortful and highly susceptible to disruption by secondary tasks (Craik et al., 1996; Mulligan, 2008). To the extent that testing effects are driven by greater elaborative processing in the retrieval than restudy condition, DA should produce greater impairments to the encoding consequences of retrieval than restudy, yielding a net decrease in the size of the testing effect.

Similarly, it is often argued that retrieval entails more effortful processing of the target stimulus than does encoding, and that this difference in effortful processing produces the testing effect (e.g., Endres & Renkl, 2015; Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007; Pyc & Rawson, 2009; Stenlund, Sundstrom, & Jonsson, 2016; also see Roediger & Butler, 2011; van den Broek et al., 2016). Consistent with this notion is the finding that more difficult retrieval conditions can enhance the size of the testing effect (e.g., Halamish & Bjork, 2011). If differences in effort produce the testing effect and enhance its size, it would likewise be expected that distraction during retrieval would impair subsequent memory to a greater degree than distraction during restudy. Consequently, views of the testing effect hinging on elaboration or effort predict a smaller testing effect under DA than under FA.

Alternatively, it may be that the encoding effects of retrieval are more similar to retrieval success itself. As noted above, retrieval success is much less affected by DA than is encoding and so it is possible that the subsequent mnemonic effects of retrieval are largely protected from the effects of distraction. This possibility seems consistent with the view of Craik, Naveh-Benjamin, and colleagues (Anderson et al., 1998; Craik et al., 1996; Naveh-Benjamin et al., 2000), who argued that retrieval is attention-demanding but obligatory, and that it takes precedence over other ongoing tasks (e.g., in dual-task situations) whereas encoding typically does not. Similarly, on the basis of experiments described below, Mulligan and Picklesimer (2016) suggested that one of the benefits of retrieval-based learning is that it is relatively protected from ongoing distraction compared to other forms of restudy. If the encoding effects of retrieval are resilient in the face of distraction, this suggests that the size of the testing effect will increase under divided attention, as the restudy condition is more greatly harmed by distraction than the retrieval condition.

Although limited, there is some prior research on divided attention and the subsequent mnemonic effects of retrieval. Dudukovic, DuBrow, and Wagner (2009), presented participants with a list of pictures in Phase 1, followed by a recognition test for some of the old items, presented under full or divided attention conditions, during Phase 2. A final recognition test revealed higher recognition accuracy for items tested during Phase 2 compared with untested items. However, the benefit of Phase 2 testing was reduced in the DA group compared with the FA group. The authors concluded that the mnemonic benefits of retrieval are minimized by divided attention (Dudukovic, DuBrow, & Wagner, 2009; see also Dudukovic, Gottshall, Cavanaugh, & Moody, 2015).

Gaspelin, Ruthruff, and Pashler (2013) assessed whether dividing attention during retrieval would increase later memory accuracy (by acting as a type of desirable difficulty). Participants studied Swahili–English word pairs in a first phase and then took a cued-recall test (either under full or divided attention) during the second phase. Both attention conditions exhibited the same amount of forgetting when assessed on a final cued-recall test, 2 days later. This led Gaspelin et al. (2013) to conclude that although dividing attention does not reduce the mnemonic effects of retrieval, it also does not act as a desirable difficulty.

Kessler et al. (2014) also examined the effects of dividing attention on retrieval and found a third conflicting result. Participants first studied pictures in an initial study phase. In the second phase, half of the pictures were tested (under full or divided attention) in a recognition memory test. A final recognition test was given on all of the pictures during the third phase. Based on the results from the final test, the authors argued that dividing attention during retrieval actually enhanced memory for those pictures.

The results of these experiments variously suggest that the encoding effects of retrieval are impaired, unaffected, or enhanced by dividing attention, suggesting a similar uncertainty regarding the effects of distraction on the testing effect, itself. However, these studies did not actually assess the testing effect. Although some early research on the testing effect compared memory for tested material with memory for materials that were not tested or represented in any way (see Roediger & Karpicke, 2006, for historical review), subsequent research argued that this comparison confounds the effects of testing with the effects of reexperience. In particular, tested material is not only retrieved it is also reexperienced whereas in an untested condition, the material is neither retrieved nor reexperienced. Based on this comparison, an apparent testing effect might reveal the actual benefit of retrieval or may merely reflect the benefit of representation. Thus, the current standard measure of the testing effect compares the retrieval condition with a restudy condition, to better equate the conditions on the reexperiencing of the material (e.g., Carpenter et al., 2008; Carrier & Pashler, 1992; Cull, 2000; Roediger & Butler, 2011; Roediger & Karpicke, 2006; Zaromb & Roediger, 2010; see Roediger & Karpicke, 2006; Rowland, 2014, for discussion).

Because Gaspelin et al. (2013) examined potential desirable difficulties during retrieval, they did not use a restudy or untested control condition. Both Dudukovic et al. (2009) and Kessler et al. (2014) used untested study items on the final test but these items had not been restudied during Phase 2. Consequently, these studies do not allow an evaluation of the testing effect as typically defined—the difference between a retrieval and a restudy condition. In addition, these studies do not allow us to determine if the effects of distraction on the encoding consequences of retrieval and study are similar, as only the retrieval condition was subjected to divided attention.

Furthermore, both Dudukovic et al. (2009) and Kessler et al. (2014) used a recognition test as the Phase 2 retrieval test, which poses further problems. First, it is not always clear what constitutes successful retrieval on a recognition test. A "hit" on a recognition test is not unambiguous evidence of successful retrieval because hits are influenced by response bias as well as actual memory accuracy.[1] Consistent with this concern, both studies exhibited substantial false alarm rates for new items, emphasizing that old responses are not always a product of successful retrieval. Recall tests are more commonly used for retrieval practice because they increase the power of the testing effect, while also providing a more direct measure of successful retrieval (Rowland, 2014). Second, recognition tests fully present an item before the retrieval attempt occurs, which means that all items are reexperienced even if not successfully recalled. This intermingling of retrieval and restudy within a purported retrieval condition makes it difficult to attribute effects of distraction to the encoding consequences of retrieval or to the encoding consequences of restudy (see Mulligan & Picklesimer, 2016, for additional detail on all of these points).

Mulligan and Picklesimer (2016) argued that assessing the effects of distraction on the mnemonic consequences of retrieval and study required additional experiments. In two experiments, word pairs were learned during Phase 1 of the experiment, and then subjected to retrieval practice (cued recall) or restudy during Phase 2, followed by a final cued-recall test (either a few minutes or 24 hr later). Attention was manipulated during Phase 2, with half the participants engaging in retrieval or restudy under full attention, and half under divided attention using a concurrent task that required classifying aurally presented digits as even or odd. The results demonstrated a testing effect, such that final recall was better in the retrieval than restudy condition, and an effect of attention, such that final recall was greater when Phase 2 was conducted under FA than DA. Most importantly, these two factors interacted such that the negative effect of DA was greater in the restudy than the retrieval condition. Alternatively stated, the testing effect increased in size when Phase 2 was conducted under DA. Consequently, the beneficial effects of testing appear to increase under distraction. Mulligan and Picklesimer (2016) interpreted this result as consistent with the idea that both retrieval success and the encoding consequences of retrieval are resilient to distraction, possibly because retrieval takes precedence over other ongoing cognitive operations whereas typical encoding processes (reflected by the restudy condition) do not (Craik et al., 1996; Naveh-Benjamin et al., 2000). Additionally, these results are at odds with the elaboration and effort accounts, which propose that the testing advantage is rooted in greater elaboration or effort, and thus this condition should be more, rather than less, susceptible to divided attention.

Mulligan and Picklesimer (2016) represents a first step in assessing the relationship between attention and the testing effect, but it does not answer all the questions on this topic. In particular, the literature on attention and memory raises two critical factors that require inquiry, one having to do with the relationship between the materials of the memory and secondary tasks, and the other having to do with the nature of retrieval's resilience to the effects of distraction.

First, research on attention and retrieval differentiates between material-general interference and material-specific interference (Fernandes & Moscovitch, 2000; see Mulligan, 2008, for review). The former refers to competition for general processing resources produced when the materials of the memory and secondary tasks are of different types, such as words for the memory test and digits for the secondary task. Material-specific interference refers to competition within the same representational system, produced when the materials of the memory and secondary tasks are drawn from the same category (e.g., both tasks use words). The latter can produce more deleterious effects on memory retrieval (e.g., Fernandes & Moscovitch, 2000, 2002, 2003).

The experiments in Mulligan and Picklesimer (2016) made use of different types of materials in the study and secondary tasks: words in the former and digits in the latter. Thus, this study focused on material general interference and found that this type of interference had substantial effects on restudy but produced quite modest effects on the mnemonic consequences of retrieval. It is unclear if the same pattern would be found in a study stressing material-specific competition. Fernandes and Moscovitch (2000, 2002, 2003) argue that material-specific interference is more likely to affect retrieval than is material-general interference. Consequently, it is natural to wonder whether material-specific competition will impact the encoding consequences of retrieval, and whether it will do so to a greater degree than occurs for restudy. A first goal of the present experiments, addressed in Experiments 1 and 2, is to examine material-specific interference, as detailed below.

The second issue has to do with the nature of retrieval's resilience to distraction, which relates importantly to secondary task performance. As noted earlier, it is typically found that secondary tasks impair encoding more than retrieval but that the costs to the secondary task are greater when paired with retrieval than encoding. This is part of the reason that Craik, Govoni, Naveh-Benjamin, and Anderson (1996; Naveh-Benjamin et al., 2000) characterized retrieval as *obligatory*, arguing that it typically takes precedence over other ongoing activities but not *automatic*, given that retrieval exacts large costs to ongoing secondary tasks. In

[1] From a dual-process perspective, one might also worry that hits may be produced by familiarity which does not entail retrieval of the original episodic context (i.e., recollection).

addition, as Craik et al. (1996) argued, manipulations varying the importance of the secondary task have large effects on encoding but not on retrieval, implying that encoding is under greater control than is retrieval. Under this view, the relatively obligatory nature of retrieval is a fundamental difference between retrieval and encoding (Craik et al., 1996; Naveh-Benjamin et al., 2000; Naveh-Benjamin, Kilb, & Fisher, 2006) and one that might drive the resilience of the encoding consequences of retrieval.

Alternatively, it could be that participants vary task priority in a strategic manner, treating the retrieval task as more important (e.g., in the context of the digit-classification task) than the restudy task. If so, then the mnemonic effects of retrieval may not be due to an obligatory retrieval process but might instead reflect controlled, strategic processes. This strategic variation in attentional allocation would be reflected in poorer secondary-task performance in the retrieval than restudy condition.

Returning to Mulligan and Picklesimer (2016), the Phase 2 results replicated the usual pattern reported in the literature. In particular, retrieval success during Phase 2 was not significantly impaired by DA but retrieval produced greater secondary task costs than did restudy. Specifically, performance on the secondary task was more accurate in the restudy than retrieval conditions—the percent correct in the secondary task was 85% and 87% for the restudy conditions of Mulligan and Picklesimer's Experiments 1 and 2, respectively, and the comparable performance was 55% and 64% in the retrieval conditions. This pattern of results was expected based on prior research and was reassuring in demonstrating that Mulligan and Picklesimer's attention manipulation behaved consistently with prior research. However, this expected result raises an ambiguity regarding the effect of DA on subsequent memory in the final recall test. In particular, the pattern of results found in these experiments could be a necessary consequence of the obligatory nature of retrieval (and representative of a fundamental difference between encoding and retrieval) or could be due to differences in prioritization between encoding and retrieval in the face of distraction. If the latter, then the elaboration and effortful accounts of the testing effect have not been properly evaluated with respect to attention. These issues are investigated in Experiments 2 and 3.

Experiment 1

Mulligan and Picklesimer (2016), using word pairs as the study materials and digits in the secondary task, found larger effects of DA in the restudy than retrieval condition (such that the size of the testing effect increased in the DA condition). This suggests that material-general interference has less influence on the encoding effects of retrieval than on the encoding consequences of additional studying. However, research on attention and memory indicates that a more stringent assessment of the effects of distraction on the testing effect requires the use of material-specific interference. Therefore, Experiment 1 used a word-based secondary task that promotes material specific interference (Fernandes & Moscovitch, 2000).

Experiment 1 used weakly related word pairs and cued recall for both retrieval practice and for the final memory test. Phase 1 consisted of an initial presentation of the word pairs. In the second phase, the word pairs were divided into four blocks and each block was presented under one of four conditions created by combining the two variables of interest: Phase 2 condition (retrieval or restudy) and attention (full or divided). Attention was divided using a word classification task previously used to examine material-specific interference (Fernandes & Moscovitch, 2000, 2002, 2003). Specifically, this task consisted of categorizing aurally presented words as referring to either man-made or natural objects. Phase 3 took place a few minutes after Phase 2 and consisted of a cued-recall test in which the first word from each pair was presented as a cue for the recall of the second word.

If the encoding consequences of retrieval are similar to the encoding effects of more typical study phases, then one might expect a marked reduction in final recall in the DA retrieval condition, perhaps commensurate with (or greater than) the expected reduction in final recall for the DA restudy condition. Indeed, elaborative and effortful processing accounts of the testing effect imply that the effects of DA should be greater in the retrieval than restudy condition, yielding a smaller testing effect under DA compared to FA. Alternatively, the encoding consequences of retrieval may be relatively resilient to the effects of distraction as retrieval success itself tends to be (Craik et al., 1996; Mulligan, 2008), suggesting that any negative effects of DA would be greater on restudy than retrieval. Under this resilience view, even if material-specific interference affects the encoding consequences of retrieval, these effects are expected to be less than the negative effects on restudy. This possibility suggests results consistent with those of Mulligan and Picklesimer (2016), in which the testing effect was larger under DA than FA.

Method

Participants. Twenty-four participants from UNC at Chapel Hill participated in exchange for course credit. The study received research ethics committee (Instructional Review Board) approval.

Design and materials. Phase 2 condition (retrieval vs. restudy) and attention during Phase 2 (full vs. divided) were manipulated within-subjects. The study list consisted of 60 weakly associated word pairs, 40 of which were drawn from Carpenter, Pashler, and Vul (2006) and the others developed to have similar properties. According to the Nelson, McEvoy, and Schreiber (2004) free association norms, the average forward strength was M = .030 (SD = 0.17; $Range$ = 0.01 to 0.07) and the average backward strength was M = .033 (SD = .021; $Range$ = 0.01 to 0.1). Examples include: *child–mother*, *office–doctor*, and *pencil–point*. Each word was four to seven letters, common (frequencies greater than 30), and of relatively high concreteness (greater than 400). The word pairs were randomly divided into four sets of 15 and assigned to the four blocks constituting Phase 2 (FA-retrieval, FA-restudy, DA-retrieval, and DA-restudy). The sets of word pairs were counterbalanced across these conditions such that each set appeared equally often in each condition. Likewise, the order of the Phase 2 blocks was counterbalanced across participants, so that each condition appeared equally often in Blocks 1 through 4.

The words used in the secondary task were common nouns that could be clearly classified into man-made or natural categories (e.g., *key*, *cat*, *pear*). None of these words were the same as words from the study list of word pairs. Overall, there were 120 of these words (60 man-made and 60 natural) and each divided attention block utilized 60 of these words, half man-made and half natural, presented in a random order.

Procedure. The experiment consisted of three phases. Phase 1 was the study phase in which participants were initially shown the 60 word pairs. Each study trial began with a 250-ms blank screen followed by the word pair for 6 s. Importantly, there was no distraction during this initial study phase. Prior to the study list, participants were told they should try to learn the word pairs for a later memory test in which the first word from each pair (the cue word) would be presented and the second (target) word should be recalled.

The second phase consisted of four blocks entailing the four possible combinations of the attention and Phase 2 conditions. Each block consisted of 15 word pairs from the initial list. In the FA-restudy block, participants were told that they would restudy a portion of the word pairs in preparation for the final test. This block contained no distractor task and simply represented the word pairs for 6 s each (preceded by a blank screen for 250 ms). For each word pair, participants were asked to read the first word to themselves and say the second word out loud (this vocal response was necessary to equate the overt response in the restudy and retrieval conditions). In the FA-retrieval block, participants were told that they would practice retrieving a portion of the word pairs in preparation for the final test. Each trial in this condition consisted of a blank screen for 250 ms followed by the full cue word and the first two letters of the target word, presented for 6 s. Participants were asked to read the first word to themselves and then recall the second word out loud. Participants were not given feedback.

The DA-restudy and DA-retrieval blocks were the same as their FA counterparts, with the addition of the secondary task. Participants were told that the secondary task and restudying or retrieving the word pairs (i.e., the memory task) were equally important. A trial in the DA blocks began with a blank screen for 250 ms followed by the presentation of a word pair (restudy block) or the first word and the first two letters of the second word (retrieval block) that stayed on the computer screen for 6,000 ms. During each trial, participants also heard a series of four words from the classification task over the headphones, at a rate of one word every 1,500 ms, with the first word being presented simultaneously with the presentation of the word pair (or cue and target stem). Participants were instructed to listen to the word and immediately press the "N" button if the word was natural (e.g., *pear*) or the "M" button if the word was man-made (e.g., *key*). If the answer was correct, no feedback was given but if the participant responded incorrectly or took too long to respond (i.e., no response after 1,500 ms), they heard a buzzer noise.

Following the second phase, participants completed 5 min of arithmetic problems on a math sheet. This was followed by the final cued-recall test (Phase 3). Each cue word was presented for 6 s followed by a blank screen for 200 ms. Participants were instructed to recall aloud the associated target word within the 6 s that the cue was presented.

Results

Cued recall. During Phase 2, the proportion of target words correctly recalled was $M = .792$ ($SD = .161$) and $M = .719$ ($SD = .144$); in the FA and DA conditions, respectively, a difference that approached traditional significance, $t(23) = 2.013$, $p = .056$.

The critical results are from the final recall test. In the context of the testing manipulation, there are two ways to measure final recall which have complementary strengths and weaknesses. First, one can assess the simple proportion of target words recalled—the unconditionalized recall score. Although straightforward, this measure produces a reexperience confound (Kuo & Hirshman, 1996; Rowland & DeLosh, 2015; Toppino & Cohen, 2009). Specifically, all the restudy items are reexperienced during Phase 2 but in the retrieval condition (without feedback as in the present experiments), only the items successfully retrieved are reexperienced. This confound advantages the restudy condition, such that the observed testing effect might underestimate the actual effects of retrieval on subsequent memory. This problem is amplified if we are interested in multiple retrieval conditions which differ in Phase 2 success rate.

A second measure conditionalizes final recall on successful Phase 2 retrieval. This measure has the virtue of eliminating the reexperience confound—now all the items in the analysis were reexperienced during Phase 2. Also, because we are interested in the encoding effects of retrieval, it seems reasonable to focus on those items that were actually retrieved. However, the conditionalized score is open to concerns about item selection effects that do not trouble the unconditionalized analysis: Unless Phase 2 retrieval is 100%, the conditionalized analysis leaves different subsets of items in the restudy and retrieval conditions (and across different retrieval conditions that differ in Phase 2 retrieval rates). Thus, neither analysis is without drawbacks. Because the analyses have complementary virtues and because there is no clear reason to prefer one over the other, both the unconditionalized and conditionalized analyses are reported. The analyses generally produce the same pattern of results and the few differences are noted.

First, the unconditionalized proportion of target words recalled on the final test (see Figure 1) was analyzed with a 2 × 2 ANOVA, using Phase 2 condition (restudy vs. retrieval) and attention during Phase 2 (full vs. divided) as within-subject factors. A main effect of attention was found, $F(1, 23) = 14.361$, $MS_e = .013$, $p = .001$, $\eta_p^2 = .384$, showing greater recall in the FA than DA condition. The main effect of Phase 2 condition was also significant, $F(1, 23) = 12.856$, $MS_e = 0.025$, $p = .002$, $\eta_p^2 = .359$, indicating

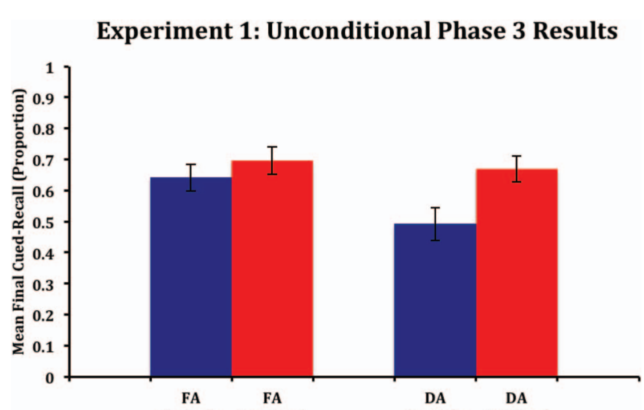

Figure 1. Experiment 1: Mean final cued-recall (±*SE*) as a function of Phase 2 condition and attention. See the online article for the color version of this figure.

higher recall in the retrieval than the restudy condition (i.e., the testing effect was found). Critically, the interaction between Phase 2 condition and attention was also significant, $F(1, 23) = 4.480$, $MS_e = .019$, $p = .038$, $\eta_p^2 = .174$, which indicates that the testing effect is larger in the DA than FA condition. Alternatively stated, the negative effect of DA was greater for the restudy than retrieval condition. Follow up analyses indicate that the effect of Phase 2 condition was significant in the DA condition, $t(23) = 3.452$, $SE = 0.252$, $p = .002$, $d = 0.259$, but only approaching significance in the FA condition, $t(23) = 1.749$, $p = .094$, indicating a significant testing effect under DA but not FA. Alternatively analyzed, the effect of attention was significant in the restudy condition, $t(23) = 4.025$, $SE = 0.037$, $p = .001$, $d = 0.642$, but not in the retrieval condition, $t(23) = 0.784$, $p = .441$, indicating that the usual negative effect of DA was found for the former but not the latter condition.

Second, recall scores on the final test (see Figure 2) were conditionalized on correct Phase 2 retrieval. These scores produced the same pattern of results as the unconditionalized analysis. The conditionalized recall proportion were $M = 0.825$ ($SD = 0.184$) and $M = 0.820$ ($SD = 0.186$) for the FA and DA retrieval conditions, respectively (the restudy conditions are unaffected by this conditionalization). These data were submitted to the same 2 (attention during Phase 2) × 2 (Phase 2 condition) ANOVA, which produced the same pattern of results. Specifically, a main effect of attention was found, $F(1, 23) = 14.844$, $MS_e = .010$, $p = .001$, $\eta_p^2 = .392$, showing that final recall was lower in the DA than FA condition. The main effect of Phase 2 condition was also significant, $F(1, 23) = 72.584$, $MS_e = .021$, $p < .001$, $\eta_p^2 = .759$, demonstrating the testing effect (i.e., higher recall in the retrieval than the restudy condition). Importantly, the interaction between Phase 2 attention and condition was still significant, $F(1, 23) = 7.339$, $MS_e = .016$, $p = .013$, $\eta_p^2 = .242$. This indicates that for the conditionalized recall data, as for the unconditionalized, the testing effect is larger in the DA than FA condition (and alternatively stated, the effect of attention is larger in the restudy than retrieval condition). Likewise, the effect of attention on the conditionalized retrieval data is nonsignificant, $t(23) = .275$, $p = .786$.

Word classification task. In the DA conditions, participants carried out the word classification task while restudying or engaging in retrieval practice. The mean proportions of accurate man-made–natural classifications were $M = .722$ ($SD = .117$) and $M = .559$ ($SD = .123$) in the DA-restudy and DA-retrieval conditions, respectively, $t(23) = 7.438$, $p < .001$, $d = 1.358$.[2]

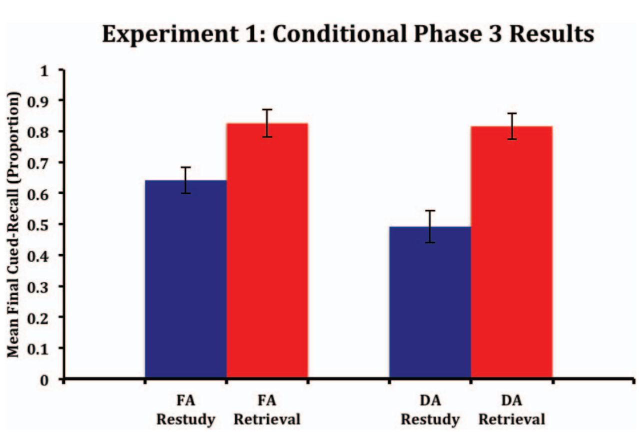

Figure 2. Experiment 1: Mean final cued-recall (±*SE*) as a function of Phase 2 condition and attention. Conditionalized on Phase 2 retrieval recall. See the online article for the color version of this figure.

Based on prior research, it was expected that Phase 2 secondary task performance would be worse in the retrieval than restudy condition. However, one might wonder whether the interaction between attention and the testing effect is driven by the large difference in performance on the secondary task. This is addressed more directly in Experiment 2 (also see Mulligan & Picklesimer, 2016), but for now we address this with an additional analysis of final recall restricted to those items for which the Phase 2 trials exhibited high accuracy on the secondary task (implying that a strong division of attention was documented on those trials). Specifically, trials with three or four (out of four) correct word classifications during a single word pair presentation were considered high accuracy. The proportion of words from high-accuracy trials that were correctly recalled on the final test was computed for each DA participant, $M = 0.476$ ($SD = 0.276$) and $M = 0.724$ ($SD = 0.278$) for the DA restudy and retrieval conditions, respectively. These data were entered into a reanalysis along with the final recall data from the FA blocks (these recall scores are unchanged given that these blocks did not include a secondary task). The results were the same as the original analyses. Most importantly, attention and Phase 2 condition interacted, $F(1, 23) = 7.418$, $MS_e = .023$, $p = .012$, $\eta_p^2 = .244$, indicating that even when the analysis is restricted to trials with high accuracy on the Phase 2 secondary task, the testing effect is larger under distraction than in the FA condition.

Discussion

The Phase 2 results are similar to those of Mulligan and Picklesimer (2016), and to much of the broader literature on attention and memory. First, performance on the secondary task was worse in the retrieval than restudy condition, consistent with the standard result finding greater secondary-task costs for retrieval than study (or restudy, in the present case; see Mulligan, 2008, for review). Second, distraction during retrieval often produces a modest reduction in retrieval success, sometimes significant and sometimes not (e.g., Craik et al., 1996). In the present case, the reduction in cued recall during Phase 2 was marginally significant.

More important for present purposes are the results from the final memory test. First, DA during Phase 2 had the expected effect on final recall in the restudy condition—encoding (or reencoding) was substantially impaired by distraction. In contrast, the effects of distraction on final recall in the retrieval condition were quite modest, resulting in a significantly larger testing effect in the DA than FA condition. This interaction remained significant when

[2] Performance in the retrieval condition may give the impression that performance in the classification task was barely above chance. This is not the case as the large majority of errors were omissions rather than incorrect responses. Restricted to trials yielding actual responses, the proportion correct on the digit classification task was $M = .860$ ($SD = .096$) and $M = .825$ ($SD = .100$), in the restudy and retrieval blocks, respectively. By this measure, word classification responses were far above chance in both conditions, $t(23)$'s > 30, p's $< .001$, although performance was still significantly better in the restudy than retrieval condition, $t(23) = 2.496$, $p = .020$, $d = 0.357$.

final recall results were conditionalized on Phase 2 memory accuracy as well as Phase 2 secondary task accuracy. In other words, the primary pattern remains even when we restrict consideration to those items that were actually subject to successful retrieval, and those items associated with high performance on the secondary task during Phase 2. This latter analysis lends support to the idea that differences in strategic allocation of attention between restudy and retrieval do not explain the significant interaction (more on this momentarily).

The present results indicate that when the memory and secondary tasks both make use of verbal materials, the results are the same as when the tasks used different materials (words vs. digits). In both cases, DA during Phase 2 produced a larger negative effect on the restudy than retrieval condition, increasing the size of the testing effect relative to the FA condition. This implies that the encoding effects of retrieval are more resilient to distraction whether the distraction is designed to induce material-specific or material-general interference (Fernandes & Moscovitch, 2000). In contrast, the results appear inconsistent with the notion that the encoding effects of retrieval produce their benefits through enhanced elaboration or effort, which implies that DA should reduce this advantage.

Experiment 2

In Experiment 1, secondary task performance was substantially better in the restudy than retrieval condition. Although this was expected based on prior research, it does raise the ambiguity described in the introduction: The difference in secondary task performance could reflect the obligatory nature of retrieval or it may reflect task prioritization differences between encoding and retrieval. If it is the latter, then this would imply that the elaboration/effort accounts and the resiliency account have not been fully evaluated. Specifically, if strategic task prioritization explains the secondary task differences, then attention was less effectively diverted in the retrieval than restudy condition. Consequently, it is still possible that the implications of the elaboration and effort accounts are correct—that the testing effect would be diminished under DA—but that a more equitable division of attention in the retrieval and restudy conditions is required to observe this outcome. Thus, for a more complete evaluation of these questions, performance on the secondary task must be made more equivalent across the Phase 2 conditions.

Experiment 2 was designed to better equate performance on the secondary task between the Phase 2 restudy and retrieval conditions. This was done by slowing the presentation of the word-classification items and by providing enhanced feedback. We believed that making the task somewhat easier (by slowing the presentation rate) would make it more likely that participants in the retrieval condition would be able to execute the secondary task at levels at least roughly equivalent to the restudy condition. This may in turn raise the concern that the secondary task would become too easy to effectively divide attention. As we shall see, this concern is unfounded: The effect of attention on final recall turned out to be quite robust.

Method

Participants. Twenty-four participants from UNC at Chapel Hill participated in exchange for course credit.

Design, materials, and procedure. The method was the same as Experiment 1 except for the following modifications. First, only 90 of the previous 120 secondary-task words were required for this experiment. Second, the Phase 2 trials in the DA condition were modified such that participants now heard three distractor words per word pair, instead of four. Specifically, during a 6,000-ms Phase 2 trial, one word was played over the headphones every 2,000 ms with the first word presented at the same time as the word pair (or cue word and target stem) was presented on the computer screen. Participants were instructed to keep track of how many of the auditory words presented in a trial were man-made (as opposed to natural). At the end of the trial, a screen appeared asking "How many of the words were man-made?" and prompted participants to enter a response (0–3) using the keyboard. If the participant responded correctly, the word "correct" was displayed on the screen; otherwise the screen displayed the word "incorrect" along with a buzzer noise. The participant then pressed the space bar to begin the next trial. This enhanced feedback, the slowed pace of the secondary task, and the change from continuous responding, were designed to decrease difficulty and better equate performance between conditions. All other aspects of Experiment 2 were identical to the first experiment.

Results

Cued recall. The proportions of target words recalled during Phase 2 retrieval were $M = .808$ ($SD = .161$) and $M = .772$ ($SD = .171$) for the FA and DA conditions, respectively. The difference in recall rates was not significant, $t(23) = 1.236$, $p = .229$.

Performance on the final recall test (see Figure 3) produced similar results as Experiment 1. The unconditionalized recall scores were submitted to a 2 × 2 ANOVA, with Phase 2 condition (restudy vs. retrieval) and Phase 2 attention (FA vs. DA) as within-subject factors. As with Experiment 1, the main effect of attention was found, $F(1, 23) = 62.575$, $MS_e = .009$, $p < .001$, $\eta_p^2 = .731$, showing greater recall in the FA than DA condition. The main effect of Phase 2 condition was also significant, $F(1, 23) = 29.754$, $MS_e = .021$, $p < .001$, $\eta_p^2 = .564$, again indicating a testing effect (i.e., higher recall in retrieval compared to restudy).

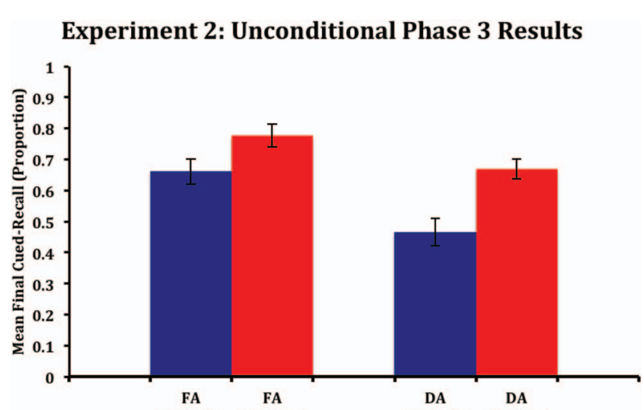

Figure 3. Experiment 2: Mean final cued-recall (±*SE*) as a function of Phase 2 condition and attention. See the online article for the color version of this figure.

Most importantly, the interaction between Phase 2 condition and attention was significant, $F(1, 23) = 6.183$, $MS_e = .007$, $p = .021$, $\eta_p^2 = .212$, indicating that the testing effect grew larger under the DA compared with the FA condition or that the negative effect of DA was smaller for the retrieval than the restudy condition.

Follow up analyses revealed that the testing effect was significant in both the FA condition, $t(23) = 4.101$, $SE = 0.028$, $p < .001$, $d = 0.618$, as well as in the DA condition, $t(23) = 5.226$, $SE = 0.039$ $p < .001$, $d = 1.068$. Additionally, the effect of attention was significant in the restudy condition, $t(23) = 6.725$, $SE = 0.029$, $p < .001$, $d = 0.936$, and also significant in the retrieval condition, $t(23) = 4.863$, $SE = .022$ $p < .001$, $d = 0.644$, suggesting that DA exhibited the usual negative effect on subsequent memory accuracy under both Phase 2 conditions.

As with Experiment 1, the final recall results (see Figure 4) were also conditionalized on correct Phase 2 retrieval, producing conditionalized recall of $M = 0.901$ ($SD = .127$) and $M = .85$ ($SD = .151$) for the FA and DA retrieval conditions, respectively (the restudy conditions are unaffected by this conditionalization). When the conditionalized data are used in place of the unconditionalized data, the critical results are unchanged. Specifically, the conditionalized recall scores were submitted to a 2 (Phase 2 condition) × 2 (attention during Phase 2) ANOVA, which produced the same pattern of results as the unconditionalized data. As before, a main effect of attention was found, $F(1, 23) = 19.875$, $MS_e = .018$, $p < .001$, $\eta_p^2 = .464$, showing that DA reduced final recall compared with FA. The main effect of Phase 2 condition was also significant, $F(1, 23) = 67.582$, $MS_e = .034$, $p < .001$, $\eta_p^2 = .746$, indicating that retrieval produced higher final recall than restudy (i.e., the testing effect). Most importantly, the interaction between attention and Phase 2 condition was still significant, $F(1, 23) = 10.616$, $MS_e = .011$, $p = .003$, $\eta_p^2 = .316$. This provides further evidence that the testing effect gets significantly larger under divided attention or that the negative effect of attention is larger under restudy compared to retrieval. Similar to Experiment 1, the effect of attention on the conditionalized recall scores is nonsignificant, $t = 1.278$, $p = .214$.

Word classification task. Experiment 2 was designed to better equate secondary task performance across the restudy and retrieval conditions than is usually the case in studies of DA and memory. The mean proportions of accurate classifications were $M = .878$ ($SD = .105$) and $M = .844$ ($SD = .102$) for the restudy and retrieval conditions, respectively, $t(23) = 1.384$, $p = .180$. That is, accuracy in the secondary task was approximately equated between the restudy and retrieval conditions. This indicates that the changes in the procedures generally succeeded in minimizing (indeed, largely eliminating) differences in secondary task performance.

In Experiment 1, an additional analysis was conducted in which the final recall data were conditionalized on high accuracy in the secondary task during Phase 2. This was done because of the substantial difference in secondary task accuracy between the retrieval and restudy conditions. Although this difference did not occur in the present experiment, this same analysis was conducted in Experiment 2 and produced the same results as the main analysis. In particular, this analysis assessed final recall performance only for word pairs associated with accurate secondary task performance during Phase 2. This produced recall scores of, $M = .479$ ($SD = .241$) and $M = .701$ ($SD = .155$) for the DA-restudy and DA-retrieval conditions, respectively. These recall scores were entered into a reanalysis with the recall scores of the FA conditions (which are unaffected by this conditionalization). Replicating the prior analyses, both main effects of attention and Phase 2 condition remained significant. Critically, the interaction was also significant, $F(1, 23) = 4.745$, $MS_e = .014$, $p = .040$, $\eta_p^2 = .171$, showing that the testing effect gets larger under distraction, even when restricting the analysis only to those items with accurate secondary task performance during Phase 2.

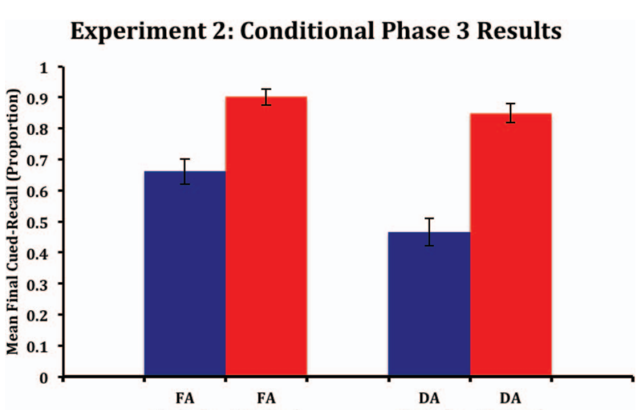

Figure 4. Experiment 2: Mean final cued-recall (±*SE*) as a function of Phase 2 condition and attention. Conditionalized on Phase 2 retrieval recall. See the online article for the color version of this figure.

Discussion

Experiment 2 replicated the critical findings of Experiment 1 and those of Mulligan and Picklesimer (2016). The final recall test showed that dividing attention during Phase 2 resulted in worse recall and that retrieval, compared with restudy, during Phase 2 boosted final recall. More critically, the interaction of attention and Phase 2 condition was significant, indicating that the testing effect significantly increases under divided attention. Additionally, this pattern of results was the same when conditionalizing on Phase 2 recall accuracy or on Phase 2 secondary task accuracy.

The goal of this experiment was to better match secondary task performance across the retrieval and restudy condition than is usual in studies of DA and memory. The results of Experiment 1 are typical, in which secondary task performance is much less accurate in a retrieval than a traditional encoding (in this case a restudy) condition. The results of Experiment 2 contrast with this usual pattern; performance on the secondary task was statistically equivalent between the two conditions. This indicates that the changes in secondary task and enhanced feedback were effective in better matching performance on the secondary task, which in turn reduces the concern that participants were not paying equivalent amounts of attention to the secondary task in the retrieval and restudy conditions. Despite this change in relative secondary task performance, the critical results on final recall persisted—the retrieval condition was less affected by DA than was the restudy condition (and thus, the testing effect was larger in the DA than FA condition). Importantly, this set of results indicates that strategic allocation of attention does not appear to account for the pattern of results reported in Experiment 1. Similar to the results obtained in Experiment 1 and by Mulligan and Picklesimer (2016), these findings are inconsistent with the elaboration and effort accounts of the testing effect which imply that DA should have a greater effect on the retrieval than restudy condition. In addition, the results converge with Experiment 1 in showing that the encoding consequences of retrieval appear to be resilient to material specific interference.

Finally, we note an aspect of the results that might seem surprising. When a verbal task is used to divide attention, memory retrieval is often significantly (and substantially) reduced (Fernandes & Moscovitch, 2000, 2002, 2003). However, in the present experiments, the word-based secondary task produced only a marginally significant reduction in Phase 2 recall (Experiment 1) or produced a nonsignificant reduction (Experiment 2). To further assess the effect of the secondary task on retrieval itself, the Phase 2 retrieval data were combined across Experiments 1 and 2. The analysis indicates that recall during Phase 2 was significantly lower in the DA, $M = .746$ ($SD = .159$), than FA, $M = .800$ ($SD = .160$) condition, $t(47) = 2.352$, $p = .023$, $d = .338$. Thus, it appears that there is an effect of the current secondary task on memory retrieval. However, the effect may seem somewhat smaller than one might expect based on the prior studies. Why? One possibility is that prior research on material-specific interference made use of free recall tests whereas the current experiments used cued recall. To our knowledge, no prior research has examined the effect of material specific interference on cued recall. Cued recall has been shown to be more resilient to DA than free recall because of the increased retrieval support in the former test (e.g., Craik et al., 1996). The difference in sensitivity to DA of free and cued recall may explain the smaller effects found during Phase 2 of the present experiments.

Experiment 3

The first two experiments examined the effects of material-specific interference on the encoding effects of retrieval by using a verbal secondary task, and found results consistent with Mulligan and Picklesimer (2016). Mulligan and Picklesimer used a digit-based secondary task (i.e., general interference) that required participants to classify digits as even or odd. However, as was the case in Experiment 1, secondary task performance differed substantially between the retrieval and restudy conditions (e.g., accuracy rates of 55% and 85%, respectively in Mulligan & Picklesimer, 2016, Experiment 1). Experiment 3 extends the analysis of Experiment 2 to the domain of material-general interference by better equating performance on the digit-classification task across the retrieval and restudy conditions.

Experiment 3 used the same materials and design as Experiment 2, aside from the change to the secondary task. In the DA conditions, participants heard three digits per word-pair over the headphones and kept track of how many of the digits were odd. At the end of the trial, participants entered the number of odd digits. Critically, the digit task was slower than in Mulligan and Picklesimer (2016) and feedback was given after every trial, in order to increase secondary task performance and minimize differences between the restudy and retrieval conditions.

Method

Participants. Twenty-four participants from UNC at Chapel Hill participated in exchange for course credit.

Design, materials, and procedure. The method of Experiment 3 was the same as Experiment 2 except for a change in the secondary task. The word classification task was replaced with the digit classification task. Thus, the modification only affected the DA Phase 2 blocks. Specifically, during these trials, participants heard three digits (drawn from the set 1–9) over the headphones during the 6,000-ms Phase 2 trial, at a rate of one digit every 2,000 ms, with the first digit being played as the word-pair was first presented. Participants were instructed to keep track of the number of odd digits presented during the trial. At the end of the trial, the screen displayed "How many of the digits were odd?" and prompted participants to enter their response (0–3) using the keyboard. The feedback for each trial was the same as in Experiment 2, as were all other aspects of the experiment.

Results

Cued recall. Retrieval during Phase 2 was significantly higher in the FA condition ($M = .872$, $SD = .104$) than the DA condition ($M = .806$, $SD = .100$), $t(23) = 3.085$, $p = .005$, $d = .647$.

A 2 (attention during Phase 2) × 2 (Phase 2 condition) within-subjects ANOVA was conducted on the unconditionalized final recall scores (see Figure 5). The main effect of attention was significant, $F(1, 23) = 79.507$, $MS_e = .009$, $p < .001$, $\eta_p^2 = .776$, showing greater recall in the FA compared with DA condition. Additionally, the main effect of Phase 2 condition, $F(1, 23) = 22.182$, $MS_e = 0.28$, $p < .001$, $\eta_p^2 = .491$, was significant, indicating that the retrieval condition produced greater recall than the restudy condition (i.e., the testing effect). However, the interaction between Phase 2 learning type and Phase 2 attention type was not significant, $F(1, 23) = .913$, $MS_e = .010$, $p = .349$.

The unconditional analyses demonstrated the main effects of attention and Phase 2 condition, as in earlier experiments, but did not show the significant interaction found in the earlier studies (and in Mulligan & Picklesimer, 2016). However, the conditionalized recall data (see Figure 6) did exhibit the interaction. When

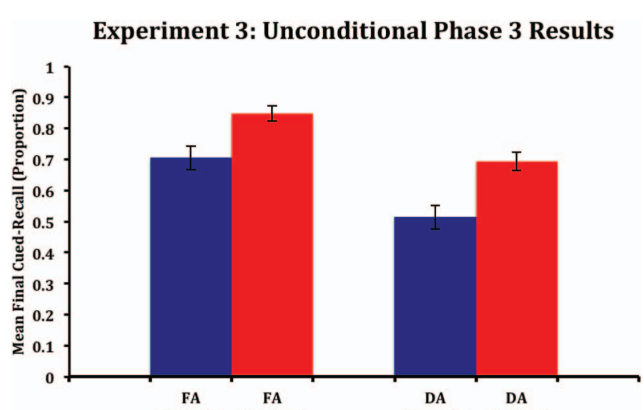

Figure 5. Experiment 3: Mean final cued-recall (±*SE*) as a function of Phase 2 condition and attention. See the online article for the color version of this figure.

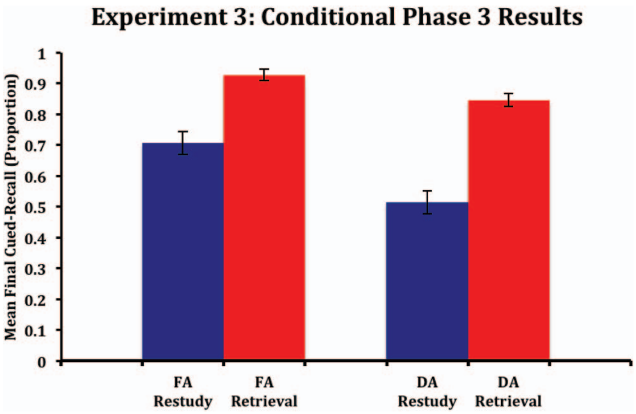

Figure 6. Experiment 3: Mean final cued-recall (±*SE*) as a function of Phase 2 condition and attention. Conditionalized on Phase 2 retrieval recall. See the online article for the color version of this figure.

final recall was conditionalized on correct Phase 2 retrieval, performance in the FA and DA retrieval conditions were $M = .927$ ($SD = .089$) and $M = .846$ ($SD = .104$), respectively. The conditionalized recall scores were submitted to a 2 (Phase 2 condition) × 2 (attention during Phase 2) ANOVA, which produced three significant effects: (a) a main effect of attention, $F(1, 23) = 47.311$, $MS_e = .009$, $p < .001$, $\eta_p^2 = .673$, where DA, as compared with FA, impaired final recall; (b) a main effect of Phase 2 condition, $F(1, 23) = 65.944$, $MS_e = .028$, $p < .001$, $\eta_p^2 = .741$, in which retrieval produced better final recall than restudy; and (c) the interaction between Phase 2 attention and condition, $F(1, 23) = 7.938$, $MS_e = .009$, $p = .01$, $\eta_p^2 = .257$. Thus, unlike the earlier experiments (and the experiments in Mulligan & Picklesimer, 2016), there is one difference in the pattern of results between the unconditionalized and conditionalized analyses: Both analyses produced significant main effects of attention and Phase 2 condition, but only the latter analysis found a significant interaction. Based on the pattern of significance, the unconditionalized data implies that the testing effect is equivalent for the FA and DA conditions, whereas the conditionalized analysis indicates that the testing effect is larger under DA than FA. However, this difference may be less stark than the pattern of significance makes it seem, as there is a numerical (but nonsignificant) trend for the same pattern in the unconditionalized data. Nevertheless, this requires additional discussion below.

Digit classification task. The mean proportion correct on the digit classification task was $M = .97$ ($SD = .039$) and $M = .93$ ($SD = .073$), for the restudy and retrieval conditions, respectively, $t(23) = 2.807$, $p = .01$, $d = .683$. Accuracy was quite high and the accuracy difference between the two conditions was much smaller relative to Mulligan and Picklesimer (2016) and Experiment 1, although secondary task performance was not fully equated.

As with the second experiment, analyses were conducted based on only those words associated with correct secondary task performance during Phase 2. This was done in terms of the unconditionalized recall scores, and in terms of the recall scores conditionalized on Phase 2 retrieval. The first analysis produced, $M = .516$ ($SD = .190$) and $M = .713$ ($SD = .135$) for the DA restudy and retrieval conditions, respectively, and the corresponding means for the second analysis were $M = .516$ ($SD = .190$) and $M = .847$ ($SD = .098$; note that the second analysis does not further affect the DA restudy condition). As before, these proportions were used along with the unchanged FA proportions to reassess the main effects and interaction of Phase 2 attention and condition. The results mirror the unconditionalized and conditionalized analysis reported above, with only the latter analysis revealing a significant interaction between Phase 2 attention and condition, $F(1, 23) = 1.9$, $p = .181$, and $F(1, 23) = 8.569$, $MS_e = .009$, $p = .008$, $\eta_p^2 = .271$, respectively.

Discussion

Several aspects of the results merit discussion. As in Experiment 2, the changes made to the secondary task greatly increased performance (relative to the implementation of this task in Mulligan & Picklesimer, 2016), indicating that in both the DA-restudy and DA-retrieval conditions, participants attended to and successfully processed the vast majority of the distractor stimuli. These changes largely (but not quite completely) eliminated the difference in secondary task performance between the two conditions. As with Experiment 2, we needn't worry that the changes to the secondary task rendered it too easy for purposes of investigating divided attention because the attention manipulation produced a substantial effect on final recall.

Turning to final recall, the experiments of Mulligan and Picklesimer (2016) also using digit-based (material-general) distraction found that DA produced a smaller effect on the retrieval than restudy condition, such that the testing effect was larger in the DA than FA condition. In the present experiment, this same numerical interaction pattern was found, but was only significant for the conditionalized analysis. Otherwise, it should be noted, the conditionalized and unconditionalized analyses produced the same pattern of results, both reporting main effects of attention and Phase 2 condition. Recall our earlier discussion of the unconditionalized and conditionalized scoring. The unconditionalized measure introduces the reexperience confound between the retrieval and restudy condition, which in the present case is further compounded by the difference in Phase 2 retrieval between the FA- and DA-retrieval conditions. The conditionalized measure eliminates the reexperience confound (between retrieval and restudy, and also between different retrieval conditions) at the potential cost of introducing item-selection differences. In the absence of perfect Phase 2 retrieval, there is no faultless measure of recall for purposes of assessing the testing effect.

Despite this ambiguity, clear conclusions can be drawn with respect to the contrast between the implications of the elaboration/effortful accounts on the one hand, and the resilience account on the other. As noted earlier, the elaboration/effortful account of the testing effect indicates that distraction should harm the retrieval condition to a greater degree than the restudy condition, producing a larger testing effect in the FA than DA condition. The results are clearly contrary to that implication: Neither conditionalized or unconditionalized analysis produces this results, and conditionalized analysis is significantly in the opposite direction, while the unconditionalized trends in the opposite direction. With regard to the resilience notion, the expectation is that the testing effect is larger in the DA than FA condition. The results are generally but not completely consistent with this prediction. Overall, the results are clearly more consistent with the resilience than elaborative/effortful view, a point we return to in the General Discussion.

General Discussion

The present experiments examined the effects of divided attention on the memory-modifying effects of retrieval. Mulligan and Picklesimer (2016) began this investigation, and found that distraction produced less effect on the encoding consequences of retrieval than restudy, resulting in a net increase in the testing effect. But a number of questions remain. A first goal was to assess whether material-specific interference (Experiments 1 and 2) produces the same pattern of results as the material-general interference instituted by Mulligan and Picklesimer (2016). A second goal was to examine the effects of distraction when steps are taken to better equate secondary task performance across the retrieval and restudy conditions (Experiments 2 and 3). Finally, an overarching goal was to assess the implications of elaborative and effortful accounts of the testing effect, on the one hand, and the resilience account of retrieval on the other.

Experiment 1 induced material-specific interference, using a word-based secondary task (Fernandes & Moscovitch, 2000, 2002, 2003), and found that the testing effect on the final recall test was significantly larger when Phase 2 occurred under divided rather than full attention. Experiment 1 also found the typical result that secondary task performance was substantially worse in the retrieval than restudy condition. Consequently, in Experiment 2, the secondary task was modified to better equate performance during Phase 2. This experiment replicated the final recall results of Experiment 1 while largely eliminating differences in secondary task performance. Experiment 3 applied to the digit classification task the same secondary-task modifications of Experiment 2, and likewise greatly reduced the typical difference in secondary task performance. In terms of final recall performance, the results were mostly consistent with prior research in showing that the testing effect increased in size in the DA compared with the FA condition, but this numerical trend was only significant for the conditionalized analysis.[3]

In the literature on attention and memory, an important distinction is made between material-general and material-specific interference (Fernandes & Moscovitch, 2000, 2002, 2003). Mulligan and Picklesimer (2016) concluded that the encoding consequences of retrieval are quite resilient to the effects of distraction, certainly more so than is restudy. Given that this study used a digit-based secondary task (i.e., material-general interference), it is natural to question whether this pattern persists with material-specific distraction. The results of Experiments 1 and 2 indicate that it does; the testing effect increases under DA with secondary tasks that promote material-specific as well as material-general interference, implying that the relative resilience of the encoding consequences of retrieval generalizes over the secondary-task materials.

With regard to secondary task performance, the results of Experiment 1 and Mulligan and Picklesimer (2016) replicate a standard pattern: Secondary task performance is better in the restudy than retrieval condition. This raises the concern that the results may be due to a differential (strategic) allocation of attention. Experiments 2 and 3 were designed to investigate this potential ambiguity by introducing modifications (e.g., the speed of the secondary task, enhanced feedback) designed to minimize the usual differences in secondary task performance between the restudy and retrieval conditions. Both experiments succeeded in this goal, greatly reducing the usual difference between restudy and retrieval. Specifically, mean accuracy on the secondary task in the restudy and retrieval conditions were 87.8% and 84.4% (Experiment 2) and 97% and 93% (Experiment 3), respectively, whereas the comparable accuracies were 72% and 56% (Experiment 1) and 85% and 55% (Mulligan & Picklesimer, 2016, Experiment 1). The success of the secondary task modifications clarifies the interpretation of the final recall data in terms of both the notion of strategic allocation of attention and the evaluation of elaborative/effortful accounts of the testing effect, as discussed next.

The present results provide a useful test of general accounts of the testing effect. The elaborative account argues that the testing effect occurs because retrieval enhances elaborative processing of the cue and to-be-retrieved target (Carpenter, 2009, 2011; Carpenter & Yeung, 2017; Rawson et al., 2015). Similarly, it has been argued that retrieval entails more effortful processing of the target item compared to restudy (Endres & Renkl, 2015; Kang, McDermott, & Roediger, 2007; McDaniel, Roediger, & McDermott, 2007; Pyc & Rawson, 2009; Stenlund, Sundstrom, & Jonsson, 2016; van den Broek et al., 2016). Both of these views imply that the testing effect should be reduced under divided attention because distraction is expected to differentially harm the condition (i.e., retrieval) more reliant on elaborative or effortful processing (Craik et al., 1996; Mulligan, 2008). Alternatively, the encoding consequences of retrieval might be resilient to distraction, in the same way that retrieval success has been conceived of as resilient (Craik et al., 1996; Mulligan & Picklesimer, 2016; Naveh-Benjamin et al., 2000). Under this possibility, final recall in the retrieval condition should be less affected by Phase 2 distraction than final recall in the restudy condition—the testing effect should increase in size under DA compared with FA.

The results of the current experiments and those of Mulligan and Picklesimer (2016) generally indicate that encoding consequences of retrieval are less affected by DA than is restudy, and that the testing effect increases under DA relative to FA. This in turn favors the view that the encoding consequences of retrieval are resilient in the face of distraction and is contrary to the pattern implied by the elaborative or effortful-processing accounts. Given the somewhat mixed results of Experiment 3, it is worth evaluating the consistency of the primary data pattern. Mulligan and Picklesimer (2016) examined the effects of divided attention on the testing effect in two experiments, each of which included two groups of subjects tested on the final recall test after either a few minutes or 24 hr. These experiments thus provide four independent evaluations of the effects of divided attention on the testing effect.

[3] Of course, the increase in the testing effect under distraction is not mediated by an absolute increase in final recall in the retrieval condition. From this perspective, the present results and those of Mulligan and Picklesimer (2016) indicate that divided attention during retrieval does not act as a "desirable difficulty" in terms of final recall performance (Gaspelin et al., 2013). Rather, distraction impaired the restudy condition to a greater degree than it impaired the retrieval condition. See Mulligan and Picklesimer (2016) for a detailed discussion.

In all four groups, the testing effect significantly increased in size under DA, a result found whether the final recall scores were unconditionalized or conditionalized on Phase 2 retrieval success. Next, the present Experiments 1 and 2 likewise found a significant interaction between attention and Phase 2 condition both for the unconditionalized and conditionalized analyses. Finally, Experiment 3 found this interaction to be significant for the conditionalized analysis but not for the unconditionalized analysis. In sum, for the conditionalized analysis, seven (of seven) data sets found a significant interaction between attention and Phase 2 condition, and for the unconditionalized analysis, six (of seven) data sets found this interaction to be significant, with the remaining data set demonstrating a numerical trend in the same direction. All of the significant interactions are in the direction implied by the resiliency notion (as is the single nonsignificant trend), and none of the data sets demonstrates a significant effect or nonsignificant trend in the direction implied by the elaborative or effortful-processing accounts. In general, the results indicate that the encoding consequences of retrieval are more robust to distraction than the sorts of encoding that occur in a more typical study (or restudy) condition.

It is also worth noting the generality of this pattern over other variables of interest. The pattern persists over shorter (a few minutes) and longer (24 hr) retention intervals (Mulligan & Picklesimer, 2016). It is found with related as well as unrelated word pairs, and whether retrieval practice uses feedback or not (Mulligan & Picklesimer, 2016). It occurs in the face of material-general or material-specific interference. The pattern is observed over higher or lower levels of secondary task performance, and whether the difference in secondary task performance between the retrieval and restudy conditions is large or minimal.

The pattern of results implies that the encoding consequences of retrieval, much like retrieval success itself, is resilient to distraction. A final issue is the nature of this resilience. Is it due to the obligatory nature of retrieval, representative of a basic difference between retrieval and encoding? Or, is the apparent resilience due to differences in task prioritization? The results of the present experiments along with those of Mulligan and Picklesimer (2016) generally argue in favor of the obligatory account and against strategic prioritization. This issue can be evaluated in two ways. First, the present Experiment 1 and both experiments in Mulligan and Picklesimer produced the standard result in which secondary task performance was much better in the restudy than retrieval condition, the exact finding that gives rise to the ambiguity regarding retrieval's resilience. For these experiments, additional analyses were conducted focusing on just those Phase 2 trials with high secondary-task performance. These analyses replicated the critical results from the primary analyses: The retrieval condition was less affected by DA than the restudy condition, producing a significantly larger testing effect under DA than FA. Second, steps were taken in Experiments 2 and 3 to better equate secondary task performance in the retrieval and restudy conditions. These experiments largely eliminated the difference in secondary task performance, and generally produced the same interaction between attention and Phase 2 condition, the only exception being the unconditionalized analysis of Experiment 3. Generally, these results argue against a strategic-prioritization account and are more consistent with the notion that retrieval is more obligatory than encoding.

Although this pattern of results seems to generalize to a number of different situations, a few interesting future directions should be explored. First, would the encoding effects of retrieval remain resilient to distraction if retrieval practice consists of free rather than cued recall? Free recall relies less on external support (i.e., no cue), is more disrupted by DA than cued-recall, and requires more strategic processing (Anderson et al., 1998; Craik et al., 1996; Moscovitch, 1992; Moscovitch, 1994). It is possible that these differences might render the encoding effects of free recall more susceptible to the negative effects of distraction. Second, another way to assess the issue of strategic allocation of attention is by using a difficult distractor task but require participants to achieve a criterion level of performance. This would allow an alternative way to equate secondary task performance in the restudy and retrieval conditions, thus allowing a converging assessment of the strategic allocation hypothesis. A third future direction involves generalizing the pattern under conditions that promote external validity. For example, more educationally relevant materials could be used along with a longer delay (e.g., a week) before the final test.

References

Anderson, N. D., Craik, F. I., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: Evidence from divided attention costs. *Psychology and Aging, 13,* 405–423. http://dx.doi.org/10.1037/0882-7974.13.3.405

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569. http://dx.doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. http://dx.doi.org/10.1037/a0024140

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13,* 826–830. http://dx.doi.org/10.3758/BF03194004

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36,* 438–448. http://dx.doi.org/10.3758/MC.36.2.438

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language, 92,* 128–141. http://dx.doi.org/10.1016/j.jml.2016.06.008

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

Chan, J. C., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 431–437. http://dx.doi.org/10.1037/0278-7393.33.2.431

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125,* 159–180. http://dx.doi.org/10.1037/0096-3445.125.2.159

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215–235. http://dx.doi.org/10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Dudukovic, N. M., Dubrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory & Cognition, 37,* 953–961. http://dx.doi.org/10.3758/MC.37.7.953

Dudukovic, N. M., Gottshall, J. L., Cavanaugh, P. A., & Moody, C. T. (2015). Diminished testing benefits in young adults with attention-deficit hyperactivity disorder. *Memory, 23,* 1264–1276. http://dx.doi.org/10.1080/09658211.2014.977921

Endres, T., & Renkl, A. (2015). Mechanisms behind the testing effect: An empirical investigation of retrieval practice in meaningful learning. *Frontiers in Psychology, 6,* 1054. http://dx.doi.org/10.3389/fpsyg.2015.01054

Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General, 129,* 155–176. http://dx.doi.org/10.1037/0096-3445.129.2.155

Fernandes, M. A., & Moscovitch, M. (2002). Factors modulating the effect of divided attention during retrieval of words. *Memory & Cognition, 30,* 731–744. http://dx.doi.org/10.3758/BF03196429

Fernandes, M. A., & Moscovitch, M. (2003). Interference effects from divided attention during retrieval in younger and older adults. *Psychology and Aging, 18,* 219–230. http://dx.doi.org/10.1037/0882-7974.18.2.219

Gaspelin, N., Ruthruff, E., & Pashler, H. (2013). Divided attention: An undesirable difficulty in memory retention. *Memory & Cognition, 41,* 978–988. http://dx.doi.org/10.3758/s13421-013-0326-5

Gates, A. I. (1917). Recitation as a factor in memorization. *Archives de Psychologie, 6,* 112–121.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 801–812. http://dx.doi.org/10.1037/a0023219

Hicks, J. L., & Marsh, R. L. (2000). Toward specifying the attentional demands of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1483–1498. http://dx.doi.org/10.1037/0278-7393.26.6.1483

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101,* 621–629. http://dx.doi.org/10.1037/a0015183

Kang, S. H., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *The European Journal of Cognitive Psychology, 19,* 528–558. http://dx.doi.org/10.1080/09541440601056620

Kessler, Y., Vandermorris, S., Gopie, N., Daros, A., Winocur, G., & Moscovitch, M. (2014). Divided attention improves delayed, but not immediate retrieval of a consolidated memory. *PLoS ONE, 9,* e91309. http://dx.doi.org/10.1371/journal.pone.0091309

Kromann, C. B., Jensen, M. L., & Ringsted, C. (2009). The effect of testing on skills learning. *Medical Education, 43,* 21–27. http://dx.doi.org/10.1111/j.1365-2923.2008.03245.x

Kuo, T. M., & Hirshman, E. (1996). Investigations of the testing effect. *The American Journal of Psychology, 109,* 451–464. http://dx.doi.org/10.2307/1423016

LeCompte, D. C., Neely, C. B., & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 472–483. http://dx.doi.org/10.1037/0278-7393.23.2.472

Lozito, J. P., & Mulligan, N. W. (2010). Exploring the role of attention during implicit memory retrieval. *Journal of Memory and Language, 63,* 387–399. http://dx.doi.org/10.1016/j.jml.2010.06.007

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *The European Journal of Cognitive Psychology, 19,* 494–513. http://dx.doi.org/10.1080/09541440701326154

McDaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14,* 200–206. http://dx.doi.org/10.3758/BF03194052

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience, 4,* 257–267. http://dx.doi.org/10.1162/jocn.1992.4.3.257

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. *Memory Systems, 224,* 369–394.

Mulligan, N. W. (2008). Attention and memory. *Learning and Memory: A Comprehensive Reference, 2,* 7–22.

Mulligan, N. W., & Lozito, J. P. (2006). An asymmetry between memory encoding and retrieval. Revelation, generation, and transfer-appropriate processing. *Psychological Science, 17,* 7–11. http://dx.doi.org/10.1111/j.1467-9280.2005.01657.x

Mulligan, N. W., & Picklesimer, M. (2016). Attention and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 938–950. http://dx.doi.org/10.1037/xlm0000227

Murdock, B. B., Jr. (1965). Effects of a subsidiary task on short-term memory. *British Journal of Psychology, 56,* 413–419. http://dx.doi.org/10.1111/j.2044-8295.1965.tb00983.x

Naveh-Benjamin, M., Craik, F. I., Perretta, J. G., & Tonev, S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *The Quarterly Journal of Experimental Psychology: Section A, 53,* 609–625. http://dx.doi.org/10.1080/713755914

Naveh-Benjamin, M., Kilb, A., & Fisher, T. (2006). Concurrent task effects on memory encoding and retrieval: Further support for an asymmetry. *Memory & Cognition, 34,* 90–101. http://dx.doi.org/10.3758/BF03193389

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36,* 402–407. http://dx.doi.org/10.3758/BF03195588

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447. http://dx.doi.org/10.1016/j.jml.2009.01.004

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335. http://dx.doi.org/10.1126/science.1191465

Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, and if so, why? Evaluating the elaborative retrieval hypothesis. *Memory & Cognition, 43,* 619–633. http://dx.doi.org/10.3758/s13421-014-0477-z

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003

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

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432–1463. http://dx.doi.org/10.1037/a0037559

Rowland, C. A., & DeLosh, E. L. (2015). Mnemonic benefits of retrieval practice at short retention intervals. *Memory, 23,* 403–419. http://dx.doi.org/10.1080/09658211.2014.889710

Stenlund, T., Sundström, A., & Jonsson, B. (2016). Effects of repeated testing on short- and long-term memory performance across different test formats. *Educational Psychology, 36,* 1710–1727. http://dx.doi.org/10.1080/01443410.2014.953037

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology, 56,* 252–257. http://dx.doi.org/10.1027/1618-3169.56.4.252

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Wirebring, L. K., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the "testing effect:" A review. *Trends in Neuroscience and Education, 5,* 52–66. http://dx.doi.org/10.1016/j.tine.2016.05.001

Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38,* 995–1008. http://dx.doi.org/10.3758/MC.38.8.995

Received January 4, 2017
Revision received March 15, 2017
Accepted April 8, 2017 ■