
The re-encoding processes of restudy and testing are equally susceptible to the impairment of divided attention

Jian Zhu*, Alicia Olechowski, Reza Habib

Department of Psychology, Southern Illinois University Carbondale, Illinois, United States of America

Abstract

After an initial study of a set of items, an intervening test of those items results in superior final memory compared with restudying those items prior to the final test. This phenomenon has been termed the testing effect. The primary purpose of the present study was to juxtapose restudy and testing in order to assess, from an attention perspective, whether the re-encoding process occurring during testing (that contributes to a superior subsequent memory, i.e., testing effect) is different from the re-encoding process during the restudy. To this end, we adopted a concurrent task to divide attention during restudy and testing. In Experiment 1, we confirmed the mnemonic impairment of divided attention during both restudy and testing on the final test. In Experiment 2, we ruled out the confounding possibility of allocation of attention between the concurrent and memory tasks by finding that allocation of attention did not affect the impairment of divided attention on final test performance. Together, our data suggest that the re-encoding component of testing is similar to the re-encoding during restudy in terms of attentional demands. The consistency of our findings with other existing theoretical accounts of testing effect and one potential methodological difference amongst the reported studies in the literature that found mixed results were discussed.

Keywords: *memory, attention, divided attention, testing effect, retrieval practice*

When verbal or pictorial materials are first studied and then either restudied or tested, subsequent retention tends to be better for tested items than for restudied items (Carpenter & DeLosh, 2005; Roediger & Karpicke, 2006b). Such mnemonic enhancement in (long-term) retention as a result of testing is referred to as the *testing effect* (see Roediger & Karpicke, 2006a; Rowland, 2014 for extensive reviews; also Delaney, Verkoeijen, & Spiguel, 2010; Pashler, Rohrer, Cepeda,

* Corresponding author:
E-mail: jerryzhu@siu.edu

& Carpenter, 2007). This phenomenon suggests that testing itself can serve as a re-encoding opportunity to support later memory performance.

An investigation of the effects of attention on the re-encoding processes during both restudy and testing could potentially provide unique insights into the cognitive mechanisms of testing effect. To set the background, previous research (see Craik, 2001; Dudukovic, 2008, Chapter 1.2, for reviews) has demonstrated that memory encoding and retrieval respond differently to the manipulation of attention. Specifically, divided attention (DA) during encoding has been shown to negatively impact subsequent remembering (Baddeley, Lewis, Eldridge, & Thomson, 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996) whereas DA during retrieval has been shown to be generally less detrimental (cf. Fernandes & Moscovitch, 2000; Hicks & Marsh, 2000).

Recent research (see Table 1) has extended DA findings from single-trial encoding and/or retrieval paradigms to multi-trial studies in order to investigate the effects of DA during restudy and retrieval on later memory. However, the findings from these studies are mixed, with the effect of DA during restudy showing a more consistent pattern of results than the effect of DA during retrieval. For example, Guez and Naveh-Benjamin (2006) found that DA, compared with full attention (FA), at encoding of word pairs that were repeated three times resulted in a significant drop in immediately cued recall performance, suggesting that DA impacts memory encoding for restudied items similarly to items that are studied only once.

With respect to studies of divided attention during retrieval, there exist three result patterns (i.e., detrimental effect of DA, no effect of DA, and beneficial effect of DA). First, Dudukovic, DuBrow, and Wagner (2009) had participants study a series of pictures under FA, and then perform a recognition test under either FA or DA. Similar to previous findings, DA had only a minimal impact on memory performance on this intervening test. It was on the final recognition test (conducted under only FA two days later) that the impairment of DA, however, emerged. In contrast, Gaspelin, Ruthruff, and Pashler (2013) found that Swahili-English word pairs tested under FA and DA were recalled equally well on the final test two days later. Kessler, Vandermorris Gopie, Daros, Winocur, and Moscovitch (2014) paradoxically observed that DA enhanced later memory (perhaps through an increase in contextual variability), if the *intervening* test was delayed to allow for memory consolidation after the initial study.

The primary purpose of the present study was to juxtapose restudy and testing in order to assess, from an attention perspective, whether the re-encoding process occurring during testing (that contributes to a superior subsequent memory, i.e., testing effect) is different from the re-encoding process during the restudy. To this end, we adopted a concurrent task to divide attention during restudy and testing. Based on the literature above, two possible kinds of results could be expected. On the one hand, DA (vs. FA) would impair final memory in both restudy

Table 1
Relevant Studies that Have Assessed the Effects of Divided Attention on Restudy and Found Inconsistent Results

Phase	Studies					
	Guez and Naveh-Benjamin (2006) ^a English word pairs	Dudukovic et al. (2009) Picture/Verbal	Gaspelin et al. (2013) ^b Swahili-English	Kessler et al. (2014) Picture	Mulligan and Picklesimer (2016) English word pairs	Current Study
Initial Study	Restudy only FA & DA	Testing only FA >= DA ^c	Testing only FA >= DA ^c	Testing only (1 d later) FA = DA	Restudy FA & DA	Restudy FA & DA
Restudy/Testing	FA > DA	—	—	FA = DA	Testing FA = DA	Testing FA = DA
Final Test	Immediate	—	—	FA < DA (1 d later)	Restudy FA > DA (1 d later)	Restudy FA > DA
	Delayed	—	FA > DA (2 d later)	FA < DA (1 d later)	Testing FA = DA (1 d later)	Testing FA > DA

Note.

^a Both initial study and restudy occurred under FA or DA in the same list in Guez & Naveh-Benjamin (2006).

^b The initial study comprised two studies plus one test under FA in Gaspelin et al. (2013).

^c Different (>) or equal (=) results were reported across multiple experiments in Dudukovic et al. (2009) and Gaspelin et al. (2013).

and testing, with the magnitude of impairment being similar in two conditions or larger in one condition than in the other. On the other hand, DA during restudy would impair final memory but DA during testing would not. To foreshow our findings, the overall results from Experiment 1 and 2 showed that DA during restudy and testing impaired final recall in both groups equivalently, thereby supporting the first outcome and suggesting that re-encoding nature of restudy and testing might be similar in terms of attentional demands. During the preparation of this manuscript, we became aware that a similar report (Mulligan & Picklesimer, 2016) was published that supports the second outcome. The authors found that their final memory performance was impaired in the restudy condition under DA, but the performance in the testing condition was not affected by DA. The inconsistency of this null effect of DA during retrieval on final memory with other existing theoretical accounts of testing effect will be discussed later. One potential methodological difference amongst these studies will also be identified.

EXPERIMENT 1

In Experiment 1, participants first studied English word pairs, then either restudied them (Restudy group) or retrieved them on an intervening cued-recall test (Testing group) under FA and DA, and finally retrieved them once again on a final free-recall test, 5 minutes later. The aim of Experiment 1 was, therefore, to investigate whether both the Restudy and Testing groups would have their final memory performance impaired under DA, and if so, whether the magnitude of the impairment would be different (smaller or larger) in the Testing group relative to the Restudy group. Although the testing effect is typically found after a time delay, it has also been shown to emerge on an immediate final test (e.g., Carpenter, 2009; Verkoeijen, Bouwmeester, & Camp, 2012). In fact, the English word pairs and test formats used in the present study were the same as those in Carpenter (2009) which previously elicited a reliable testing effect after a brief 5-min test-delay.

METHOD

Participants

Sixty college students (aged 18-30 years) who were native English speakers participated in Experiment 1. Participants either received course credit or a \$10 gift card as compensation. All procedures were approved by the institutional Human Subjects Committee. Half of the participants were randomly assigned to the Restudy group; the other half to the Testing group.

Materials

Forty-eight target words were chosen from the set used in Carpenter (2009), 12 of which were randomly selected for each participant for practice purpose and the remaining 36 served as the to-be-studied set. Each target word was equally likely paired with a strongly- or weakly-associated cue word and then evenly assigned to the full and divided attention conditions in a randomized fashion for each participant.

Auditory stimuli that were used in the concurrent task to create the DA condition were provided courtesy of Dr. Nicole M. Dudukovic and consisted of two 1500-msec distinct sound tones (16 bits, Mono, 44100 Hz; Dudukovic et al., 2009). The tones were presented over a speaker to participants.

Procedure

Participants were tested individually. After participants signed a consent form, they were shown 48 word pairs (e.g., tea-coffee) in the center of a computer screen. They were instructed to study each word pair by rating its relatedness for a later memory test but they were not informed of the format of the later test. Each word pair was presented for 3000 ms with a 500 ms inter-stimulus interval (ISI).

Following the initial study phase, participants were first familiarized with the concurrent task and then performed it alone for 105 s. This provided a measure of baseline performance on the concurrent task for comparison to performance in the following DA condition. The concurrent task was presented auditorily to avoid modality-specific interference with the visual word-pair tasks. It consisted of two types of tones played sequentially without gap. From one presentation to the next, one tone may have switched to the other tone or may have remained the same; this was determined in a random fashion. Participants were required to press the space bar only when they detected a change from one tone to the other. Both accuracy and reaction times were recorded.

Immediately following the concurrent task baseline, participants practiced either the Restudy or Testing task, under both FA (i.e., alone) and DA (i.e., in the presence of the concurrent task). For the Restudy practice under FA, each participant saw six random word pairs from the study phase (one pair at a time) and was instructed to read aloud the word pairs to the experimenter who then recorded the responses. For the Restudy practice under DA, the participant saw a different set of six words and was instructed to read aloud the word pairs to the experimenter while at the same time listening for and responding to tone switches by pressing the space bar. The Testing practice under FA and DA were similar to the Restudy practice except that the target word on the right was replaced with a question mark and participants were instructed to read aloud the cue word first and then recall

aloud the target word to the experimenter who recorded the answer but did not provide any feedback indicating whether it was correct. During both practice tasks, each word pair was presented for 3500 ms with 0 ms ISI. Under DA, participants were instructed to treat both tasks (the concurrent task and either the Restudy or Testing task) as equally important and that they should respond as accurately/quickly as possible on both tasks.

Next, half of the remaining 36 word pairs were restudied or tested while the participants simultaneously performed the concurrent task (i.e., the DA condition); the other half were restudied/tested under FA. The FA and DA conditions were the same as those during the practice phase and alternated in blocks of six word pairs, for a total of six blocks. At the beginning of each block, participants were informed whether they would perform the memory task (restudy/testing) under FA or DA. The presentation order of FA and DA was randomized and counterbalanced across participants in the Restudy and Testing groups respectively.

After a 5-min distractor phase during which participants solved simple mathematic addition and multiplication problems, participants were asked to rate how they allocated their attention between the concurrent task and the memory (restudy/testing) task using a 5-point Likert scale (1 = 0% to the memory task, 2 = 25% to the memory task, 3 = 50% to the memory task, 4 = 75% to the memory task, 5 = 100% to the memory task). Following the rating questionnaire, participants received a final free-recall test under FA for 10 min in which they were instructed to recall aloud as many previously presented target words as they could to the experimenter who recorded their answers. Any words other than the 36 target ones were considered incorrect responses. After the completion of the experiment, participants were fully debriefed and dismissed. The entire experimental session lasted approximately 40 min.

Analytic approach

The design of Experiment 1 consisted of a 2 Group (Restudy vs. Testing) x 2 Attention (FA vs. DA) mixed design. Group was manipulated between participants whereas Attention was manipulated within participants. For the concurrent task used to create the DA condition, the dependent variables were accuracy (hits – false alarms) and median reaction times to the correct (“hit” only) detection of tone switches (see below). The number of words correctly recalled served as the dependent variable during on the intervening cued-recall task (Testing group) and on the final free-recall test (both groups).

RESULTS AND DISCUSSION

Concurrent task performance

Accuracy (hits – false alarms) on the concurrent task was subjected to a 2 Phase (Baseline vs. DA) x 2 Group (Restudy vs. Testing) mixed-design ANOVA. Accuracy dropped when the concurrent task was performed during DA ($M = .50$, $SE = .03$) compared with when it was performed alone during Baseline ($M = .79$, $SE = .03$) regardless of Group, $F(1, 21) = 46.05$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .39$. Accuracy was higher in the Restudy group ($M = .76$, $SE = .04$) than the Testing group ($M = .62$, $SE = .04$) regardless of the Phase, $F(1, 58) = 7.23$, $MSE = 0.04$, $p = .009$, $\eta_p^2 = .11$. Importantly, the interaction of Phase and Group was also significant, $F(1, 58) = 28.79$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .33$, which was further analyzed using a simple main effects analysis. As shown in Table 2, during Baseline, the Restudy and Testing groups performed similarly, $F(1, 58) < 1$. However, during DA, the Restudy group had better accuracy than the Testing group, $F(1, 58) = 23.55$, $MSE = 0.05$, $p < .001$, $\eta^2 = .29$.

Table 2

The Accuracy and Reaction Time on the Concurrent Task in Experiment 1

Phase	Accuracy		Reaction Time (ms)	
	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI
Baseline				
Restudy	.77 (.23)	[.68, .85]	474 (123)	[434, 514]
Testing	.76 (.23)	[.68, .85]	450 (77)	[409, 490]
DA				
Restudy	.75 (.21)	[.67, .83]	515 (95)	[470, 560]
Testing	.47 (.23)	[.39, .56]	579 (155)	[535, 624]

Note. Each group (Restudy and Testing) had 30 participants. Accuracy was the result of hits minus false alarms. The means of reaction time were averaged median reaction time to “hit” reactions across participants. Baseline refers to the concurrent baseline phase when the concurrent task was performed alone, whereas DA refers to the time when the concurrent task was performed together with the memory task during the restudy/testing phase. See the text for statistical significance. CI = confidence interval.

Similar to the analysis of accuracy, median reaction time to “hit” responses on the concurrent task was subjected to a 2 Phase (Baseline vs. DA) x 2 Group (Restudy vs. Testing) mixed-design ANOVA. Reaction time increased when the concurrent task was performed during DA ($M = 547$ ms, $SE = 16$ ms) compared to when it was performed alone during Baseline ($M = 461$ ms, $SE = 14$ ms) regardless of Group, $F(1, 58) = 24.64$, $MSE = 8851.02$, $p < .001$, $\eta_p^2 = .30$. Reaction times did not differ between the Restudy group ($M = 494$ ms, $SE = 17$ ms) and the Testing

group ($M = 514$ ms, $SE = 17$ ms) regardless of Phase, $F(1, 58) < 1$. Importantly, the interaction of Phase and Group was significant, $F(1, 58) = 6.63$, $MSE = 8851.02$, $p = .01$, $\eta_p^2 = .10$. A simple main effects analysis revealed that reaction times did not differ between the Restudy and Testing groups during Baseline, $F(1, 58) < 1$. During DA, however, the Restudy group responded faster than the Testing group, $F(1, 58) = 6.20$, $MSE = 17032.51$, $p = .02$, $\eta^2 = .10$ (see Table 2).

Intervening memory

A paired samples t -test revealed that in the Testing group, performance on the intervening cued recall test under FA ($N = 30$, $M = .72$, $SD = .16$) was *marginally* better than under DA ($M = .68$, $SD = .17$), $t(29) = 1.95$, $p = .06$, 95% CI [0, .08], $d = 0.36$.

Final memory

The memory performance on the final free-recall test was subjected to a 2 Group (Restudy vs. Testing) \times 2 Attention (FA vs. DA) mixed-design ANOVA. As shown in Figure 1, the Testing group ($M = .33$, $SE = .02$) recalled more words than the Restudy group ($M = .26$, $SE = .02$) regardless of Attention, $F(1, 58) = 6.76$, $MSE = 0.01$, $p = .01$, $\eta_p^2 = .10$. Participants also recalled more words when the intervening memory task was performed under FA ($M = .32$, $SE = .02$) than DA ($M = .27$, $SE = .02$) regardless of Group, $F(1, 58) = 5.47$, $MSE = 0.02$, $p = .02$, $\eta_p^2 = .09$. However, the interaction of Group and Attention was not significant, $F(1, 58) < 1$.

Two further analyses were carried out. First, an analysis of final memory performance conditionalized on correct recall on the intervening test (that is, only words that were successfully retrieved on the intervening test for the Testing group were examined on the final test; the Restudy Group was not affected by this conditionalization) produced the same outcome as above. Overall, the Testing group ($M = .41$, $SE = .02$) recalled more words than the Restudy group ($M = .26$, $SE = .02$), $F(1, 58) = 33.32$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = .37$; words that were restudied/tested previously under FA ($M = .36$, $SE = .02$) were better recalled than those previously under DA ($M = .31$, $SE = .02$), $F(1, 58) = 4.27$, $MSE = 0.02$, $p = .04$, $\eta_p^2 = .07$. Group did not interact with Attention, $F(1, 58) < 1$.

Second, a final memory difference was calculated between FA and DA for each participant to represent the mnemonic impairment of DA (vs. FA) during Restudy and Testing. The difference scores were then subjected to an independent t -test. Again, the Restudy group ($M = -.07$, $SE = .03$) did not have a different impairment from DA compared with the Testing group ($M = -.04$, $SE = .03$) on the final test, $t(58) = -.68$, $p = .50$, 95% CI [-.12, .06], $d = -.18$.

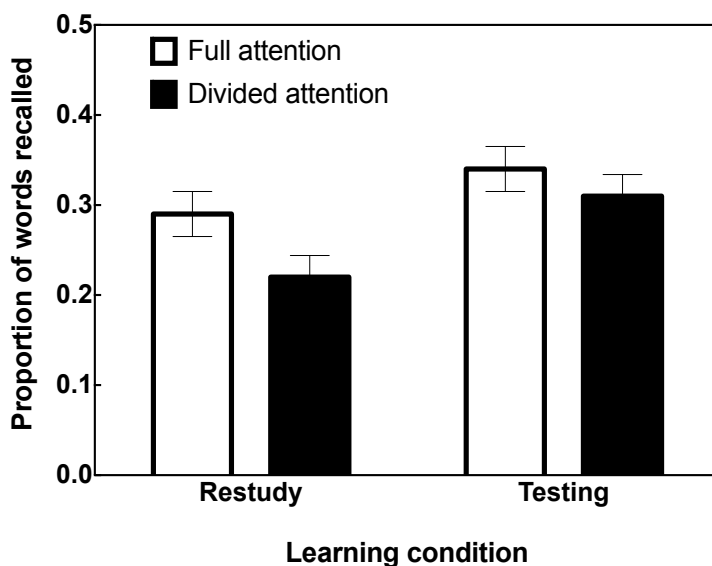


Figure 1. Memory on the final free recall test (5 min later) for Restudy and Testing groups under both FA (full attention) and DA (divided attention) in Experiment 1. Overall the Testing group recalled more words than the Restudy group, $p = .01$, and participants recalled more words when the intervening memory task was performed under FA than DA, $p = .02$. Critically, the interaction of Group and Attention was not significant $F(1, 58) < 1$. Error bars represent standard errors.

Allocation of attention self-report

A Mann-Whitney U test was conducted to examine whether participants in the Restudy group allocated attention differently from those in the Testing group. Two participants in the Testing group failed to provide responses on the questionnaire and were excluded from this analysis. The analysis indicated that participants in the Restudy group allocated less attention to the intervening memory task than participants in the Testing group, $z = -2.34$, $p = .02$.

To summarize, on the intervening cued-recall test (the Testing group only), participants recalled marginally more target words under FA than DA. On the final free-recall test, participants in the Testing group recalled more words than those in the Restudy group, replicating previous testing effect findings (Carpenter, 2009). Importantly, final memory recall was significantly better for word pairs tested earlier under FA than DA, and this DA impairment was comparable for both Restudy and Testing groups.

One tentative interpretation of the above results is that focused attention is equally important to both restudy and testing as learning opportunities for

subsequent memory. However, performance on the concurrent task during DA was poorer (lower accuracy, longer reaction times) in the Testing group than in the Restudy group, a finding similar to that observed by Mulligan and Picklesimer (2016). Additionally, participants themselves in the Testing group indicated that they allocated more of their attention to the intervening cued recall task than those in the Restudy group allocated to the restudy task. Therefore, a group difference in the allocation of attention between tasks could have possibly confounded the final memory result. In other words, it remains unclear whether a change in the allocation of attention would modulate the impairment of DA during restudy and testing.

EXPERIMENT 2

The goal of Experiment 2 was to directly manipulate the allocation of attention between the intervening memory task and concurrent task to determine its confounding possibility. Therefore, Experiment 2 focused exclusively on the DA condition (i.e., the FA condition was not included during the restudy or intervening test). If the impairment of DA is found not to be susceptible to changes in the allocation of attention (e.g., less attention paid to the concurrent task does not lead to less impairment of DA during restudy and testing), then the group difference in the allocation of attention unlikely confounded the final memory result in Experiment 1, and the comparable detrimental effect of DA (vs. FA) for both Restudy and Testing groups on final memory in Experiment 1 would be further corroborated.

Craik and colleagues (1996) provided an approach to implementing the differential allocation of attention. In one condition of the study, participants were told a memory task was more important than a concurrent task; in a second condition both tasks were emphasized equally; in the third condition participants were told the concurrent task was more important than the intervening memory task. Their data showed that higher priority assigned to the memory task resulted in poorer performance on the concurrent task, and vice versa. The authors concluded that the allocation of attention between both tasks was somewhat under the control of participants. Consequently, we adopted the same strategy and asked participants to proportionally allocate their attention during the DA condition in Experiment 2.

METHOD

Participants

Fifty college students (aged between 18-30 years) participated in Experiment 2. Their participation was compensated by extra course credit or a \$10 gift card. All procedures were approved by the institutional Human Subjects Committee. Half of the participants were randomly assigned to the restudy condition; the other half were assigned to the testing condition.

Materials

The materials for the memory task and auditory stimuli for the concurrent task were the same as those used in Experiment 1.

Procedure

Participants were told that they would be participating in a multitasking experiment but were not informed about either the intervening or final memory test. The procedure for the initial study phase was the same as Experiment 1. During the concurrent task baseline that followed the study phase, participants practiced the concurrent task alone and then together with their respective intervening memory task (either restudy or cued recall) for which 12 out of 48 word pairs were selected; unlike in Experiment 1, there was no practice for the intervening memory task under full attention.

The restudy/testing phase was also similar to that in Experiment 1 except for the following changes. The remaining 36 word pairs were split into three blocks of 12 pairs. In the spirit of Craik et al. (1996), participants were instructed that (a) for some blocks, the concurrent task was much more important than the intervening memory task and therefore they should pay 80% of their attention to the concurrent task and 20% to the intervening memory task; (b) for some blocks, the two tasks were equally important, and therefore they should perform their best on both tasks (i.e., 50%/50% allocation of attention); and (c) for other blocks, the intervening memory task was much more important than the concurrent task, and therefore they should pay 20% of their attention to the concurrent task and 80% to the memory task. The presentation order of these three conditions was randomized and counterbalanced for participants within either the Restudy or Testing groups. Strongly- and weakly-associated word pairs were evenly assigned to each condition. Both Restudy and Testing occurred under DA only.

Following the distractor task which was the same as that in Experiment 1, participants were asked to rate how they actually allocated their attention across all blocks (1 = always paid more attention to the concurrent task, 2 = always more

attention to the memory task, 3 = always equal attention to both tasks, 4 = followed the instructions, 5 = did something else and should tell the experimenter).

Participants then performed a final free-recall test similar to that in Experiment 1 under FA for 5 min. (In Experiment 1, participants were found to recall very few words during the second half of the 10 min of final recall time. Therefore, the time in Experiment 2 was scaled down to 5 min). Following the end of Experiment 2, participants were asked to indicate whether they had expected a final memory test earlier using a 5-point Likert scale (1 = confidently unexpected, 2 = somewhat unexpected, 3 = not sure, 4 = somewhat expected, 5 = confidently expected). After the completion of the experiment, participants were fully debriefed. The entire session lasted approximately 40 min.

Analytic approach

The design of this experiment consisted of a 2 Group (Restudy vs. Testing) x 3 DA (more attention to the concurrent task [Concurrent], equal attention to both tasks [Equal], more attention to the intervening memory task [Memory]) mixed design. Group was manipulated between participants whereas DA was manipulated within participants. For the concurrent task used to create the DA condition, the dependent variables were accuracy (hits – false alarms) and median reaction times to the correct (“hit” only) detection of tone switches. The number of words correctly recalled served as the dependent variable during the intervening cued-recall task (Testing group) and on the final free-recall test (both groups).

RESULTS AND DISCUSSION

One participant in the Restudy group and another participant in the Testing group misunderstood the task instructions; their data were, therefore, not included in the analyses.

Concurrent task performance

Because there were three types of allocation of attention under DA, an a priori contrast between Baseline and the “averaged” DA across all three conditions was constructed and compared between groups. Accuracy (hits – false alarms) on the concurrent task differed between the Restudy and Testing groups during Baseline and under DA (Table 3, Baseline vs. Averaged DA), $F(1, 46) = 38.08$, $MSE = 0.35$, $p < .001$, $\eta_p^2 = .45$. A simple main effect analysis performed separately for Baseline and averaged DA conditions showed that the Restudy group did not perform differently from the Testing group during Baseline, $F(1, 46) < 1$, but the Restudy

group was overall more accurate than the Testing group under DA, $F(1, 46) = 37.97$, $MSE = 0.28$, $p < .001$, $\eta_p^2 = .45$.

Table 3

The Accuracy and Reaction Time on the Concurrent Task in the DA Condition in Experiment 2

	Accuracy		RT (ms)	
	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI
Baseline				
Restudy	.67 (.27)	[.56, .79]	512 (152)	[456, 569]
Testing	.71 (.19)	[.63, .79]	495 (122)	[438, 551]
Concurrent DA				
Restudy	.76 (.22)	[.67, .86]	555 (138)	[487, 623]
Testing	.47 (.17)	[.40, .54]	665 (188)	[598, 733]
Equal DA				
Restudy	.70 (.23)	[.60, .80]	575 (169)	[510, 639]
Testing	.40 (.24)	[.30, .50]	667 (144)	[602, 731]
Memory DA				
Restudy	.66 (.24)	[.56, .76]	599 (156)	[526, 671]
Testing	.31 (.25)	[.20, .41]	606 (194)	[534, 678]
Averaged DA				
Restudy	.71 (.23)	[.63, .78]	576 (154)	[523, 629]
Testing	.39 (.22)	[.32, .47]	648 (175)	[593, 699]

Note. Each group (Restudy and Testing) had 24 participants. Accuracy was the result of hits minus false alarms. The means of reaction time were averaged median reaction time to “hit” reactions across participants. Baseline refers to the concurrent baseline phase when the concurrent task was performed alone, whereas DA refers to the time when the concurrent task was performed together with the memory task during the restudy/testing phase. The averaged DA is the mathematical mean of the three DA (Concurrent, Equal, Memory) conditions. See the text for statistical significance. CI = confidence interval.

Next, accuracy on the concurrent task was subjected to a 3 DA (Concurrent, Equal vs. Memory) x 2 Group (Restudy vs. Testing) mixed design ANOVA. The Restudy group ($M = .71$, $SE = .04$) had higher accuracy than the Testing group regardless of types of DA ($M = .39$, $SE = .04$), $F(1, 46) = 37.96$, $MSE = 0.03$, $p < .001$, $\eta_p^2 = .45$. Accuracy was different among the three types of DA regardless of Group, $F(2, 92) = 6.78$, $MSE = 0.03$, $p = .002$, $\eta_p^2 = .13$. The interaction of DA and Group was not significant, $F(2, 92) < 1$ (Table 3). A post-hoc test of the main effect of DA using Bonferroni adjustment for multiple comparisons showed that, among the three types of DA, only the Concurrent DA ($M = .62$, $SE = .03$) had higher accuracy than the Memory DA ($M = .48$, $SE = .04$), $p = .003$; the Equal DA ($M = .55$, $SE = .03$) did not significantly differ in accuracy from the other two levels of this variable, $ps > .23$. However, there was a significant

linear trend for types of DA (Concurrent > Equal > Memory), $F(1, 46) = 12.47$, $MSE = 0.03$, $p = .001$, $\eta_p^2 = .21$.

A similar a priori contrast between Baseline and the averaged DA was constructed and compared between groups to analyze median reaction time to “hit” reactions on the concurrent task. Reaction time differed between the Restudy and Testing groups during Baseline and under DA (Table 3, Baseline vs. Averaged DA), $F(1, 46) = 5.12$, $MSE = 161598.63$, $p = .03$, $\eta_p^2 = .10$. A simple main effect analysis performed separately for Baseline and DA showed that the Restudy group did not have a different reaction time from the Testing group during Baseline, $F(1, 46) < 1$, but the Restudy group overall tended to respond faster than the Testing group under DA, $F(1, 46) = 3.52$, $MSE = 150008.97$, $p = .07$, $\eta_p^2 = .10$.

Next, median reaction times on the concurrent task were subjected to a 3 DA (Concurrent, Equal vs. Memory) x 2 Group (Restudy vs. Testing) mixed design ANOVA. Reaction times were not different among the three types of DA regardless of Group, $F(2, 92) < 1$, but the Restudy group overall tended to respond faster than the Testing group under DA, $F(1, 46) = 3.52$, $MSE = 16667.66$, $p = .07$, $\eta_p^2 = .10$. The interaction of DA and Group was not significant, $F(2, 92) = 2.21$, $MSE = 16327.70$, $p = .12$, $\eta_p^2 = .05$ (Table 3).

Intervening memory

In the Testing group, participants recalled 54% of the target words ($SD = .19$) on the cued-recall test under Concurrent DA, 60% ($SD = .22$) under Equal DA, and 63% ($SD = .18$) under Memory DA. A repeated measures ANOVA showed that memories on the intervening cued-recall test did not differ as a function of types of DA, $F(2, 46) = 1.90$, $MSE = 0.03$, $p = .16$, $\eta_p^2 = .08$.

Final memory

The memory performance on the final free-recall test was subjected to a 2 Group (Restudy vs. Testing) x 3 DA (Concurrent, Equal vs. Memory) mixed-design ANOVA. As shown in Figure 2, the Testing group ($M = .21$, $SE = .02$) recalled more words than the restudy group ($M = .13$, $SE = .02$) regardless of types of DA, $F(1, 46) = 12.60$, $MSE = 0.02$, $p = .001$, $\eta_p^2 = .22$. Regardless of Group, memory performance did not differ as a function of types of DA, $F(2, 92) = 1.54$, $MSE = 0.01$, $p = .22$, $\eta_p^2 = .03$. The interaction of Group and DA was also not significant, $F(2, 92) = 1.17$, $MSE = 0.01$, $p = .32$, $\eta_p^2 = .03$.

These results remained unchanged when final memory performance was conditionalized on successfully retrieved words on the intervening test. Overall, the Testing group ($M = .32$, $SE = .02$) recalled more words than the restudy group ($M = .13$, $SE = .02$), $F(1, 46) = 30.78$, $MSE = 0.04$, $p < .001$, $\eta_p^2 = .40$; types of DA did not have different effects on memory, $F(2, 92) = 1.52$, $MSE = 0.02$, $p = .22$,

$\eta_p^2 = .03$. Group also did not interact with types of DA, $F(2, 92) = 1.45$, $MSE = 0.02$, $p = .24$, $\eta_p^2 = .03$.

Allocation of attention self-report

A Fisher's exact test showed that participants in the Restudy group did not allocate attention between the concurrent and memory tasks differently from those in the Testing group, $\chi^2 = 2.65$, $p = .66$.

Expectation of a final memory test

Finally, 17 (71%) of the participants in the Restudy group and 19 (79%) in the Testing group did not expect a final memory test (choosing confidently or somewhat unexpected). A Mann-Whitney U test showed that the expectation of a final memory test did not differ between the Restudy group and the Testing group, $z = -0.28$, $p = .84$. The Restudy group had an average rank of 24.00 whereas the Testing group had an average rank of 25.00.

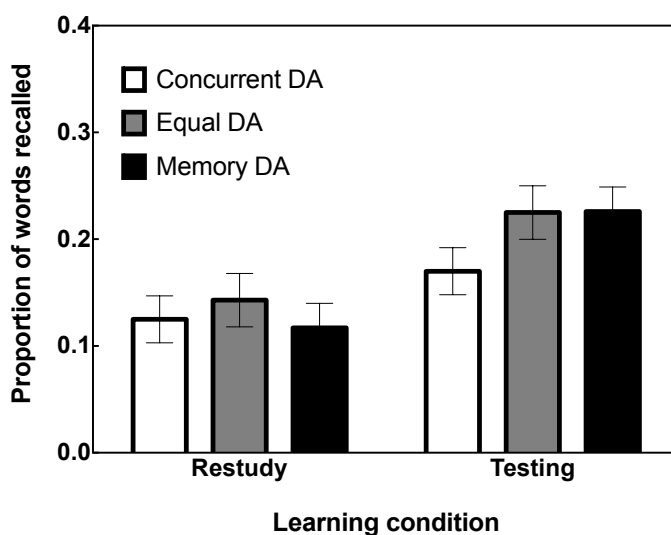


Figure 2. Memory on the final free recall test (5 min later) for Restudy and Testing groups under different types of DA (i.e. paying more attention to the concurrent task, equal attention to both tasks and more attention to the memory task) in Experiment 2. Overall, the Testing group recalled more words than the restudy group, $p = .001$. However, final memory performance did not differ as a function of types of DA or the interaction between types of DA with Group, $ps > .03$. Error bars represent standard errors.

To summarize, Experiment 2 served the purpose of examining the impact of differential allocation of attention during restudy/testing on final memory. It was accomplished by manipulating the relative priority/importance of the concurrent and memory tasks (i.e., Concurrent, Equal, and Memory DA). This manipulation was somewhat successful in that the accuracy of both the Restudy and Testing groups on the concurrent task in the DA conditions showed a linear trend for types of DA (Concurrent > Equal > Memory).

On the intervening cued-recall test (the Testing group only), the three types of DA did not produce differential memory performance. On the final test, the Testing group recalled more words on the final free-recall test than the Restudy group. However, types of DA did not have an impact on final memory either as a main effect nor as an interaction with Group, thus implying that the comparable impairment of DA during restudy and testing found in Experiment 1 was not confounded by the group difference in the allocation of attention. Because the majority of the participants did not expect either an intervening or final memory test, the overall memory performance was slightly lower on the intervening and final tests in Experiment 2 than that in Experiment 1 (see Pu & Tse, 2014 for a similar finding).

GENERAL DISCUSSION

Elucidating the cognitive mechanisms underlying the mnemonic advantages of testing (vs. restudy) in relation to focused (vs. divided) attention is of interest to both memory researchers and educational practitioners. In both Experiment 1 and 2, we replicated the testing effect, irrespective of attention conditions, with an immediate final test as previously reported in Carpenter (2009). Moreover, in Experiment 1, DA (vs. FA) during restudy/testing was shown to impair final memory to a comparable extent. One possible concern with the interpretation of these results was that attention was allocated differentially to the memory task as a function of Group (Restudy vs. Testing). In Experiment 2, it was found that directly manipulating the allocation of attention between the concurrent and intervening memory tasks did not impact memory performance on the final test in both groups. Taken together, our data suggest that the re-encoding processes during restudy and testing are equally susceptible to the impairment of divided attention.

The finding in Experiment 1 that DA during restudy impaired subsequent final free-recall memory is consistent with findings by Guez and Naveh-Benjamin (2006) which demonstrated that DA (vs. FA) at repeated encoding caused a significant drop in immediately cued recall. It is also consistent with the finding of worse final memory performance for word pairs previously restudied under DA (vs. FA, i.e., the restudy condition) in Mulligan and Picklesimer (2016). Our Experiment 2 further showed that this impairment of DA during restudy on final

memory did not vary with the differential allocation of DA (i.e., how attention was allocated between the restudy and concurrent tasks), suggesting that the impairment on memory during encoding resulting from DA is likely to be of a more *qualitative* than quantitative nature.

On the intervening test (the Testing group only) participants in Experiment 1 recalled slightly more target words under FA (.72) than under DA (.68). Dudukovic et al. (2009), Gaspelin et al. (2013), and Mulligan and Picklesimer (2016) all found similar results on their intervening tests where memory retrieval was equal or marginally better for items tested under FA vs. DA. These results are consistent with findings in the literature that demonstrate only a minimal (e.g., Baddeley et al., 1984; Craik et al., 1996), but occasionally (borderline) significant, impairment of DA during memory retrieval (Hicks & Marsh, 2000).

Performance on the concurrent task during DA was found to be poorer (lower accuracy, longer reaction times) in the Testing group than in the Restudy group in Experiment 1, indicating that attention was allocated differentially to the memory task as a function of Group, and thus raising the possibility of attention allocation as a confounding factor. This same pattern of results on the concurrent task was also found in both experiments by Mulligan and Picklesimer (2016). In Experiment 2, we directly manipulated attention allocation between the intervening memory tasks and the concurrent task through different task priorities. As a result of this manipulation, accuracy (but not RT) increased on the concurrent task in both the restudy and testing groups as the memory tasks were less prioritized, indicating that participants were (somewhat) able to differentially allocate attention between the concurrent and memory tasks. Despite the differential allocation of attention, however, performance on the final memory test did not differ as a function of the attention manipulation. This finding supports our conclusion that DA comparably affects the re-encoding process during both restudy and testing. As an interesting side note, Craik et al. (1996, Experiment 3) found that RT (but not accuracy) on the concurrent task varied as a function of types of DA during both memory encoding and retrieval, thus indicating that a speed-accuracy trade-off on the concurrent task may not always be observed.

The most important result in the current study concerns the finding of a comparable effect of DA on final memory performance across the restudy and testing groups. This result, consistent with the detrimental effect found in Dudukovic et al. (2009), implies that the re-encoding process during testing is similar to the re-encoding process during restudy in terms of susceptibility to the division of attention. We agree with Dudukovic et al.'s (2009) reasoning that, although retrieval itself is protected under DA and can be successful during the intervening testing, the re-encoding process intrinsic to that retrieval phase is not as resilient to the mnemonic impairment of DA for subsequent memory. This conclusion is consistent with the general notion that (re-)encoding is attention dependent and sensitive to distraction. Conversely, the null effect of DA during

retrieval on final memory reported in Gaspelin et al. (2013) and Mulligan and Picklesimer (2016) contradicts with other theoretical accounts of the testing effect. For instance, the effortful processes during testing such as the mediation/elaboration between the cue-target (Carpenter, 2004, 2011; Pyc & Rawson, 2010) and information organization/integration (Carpenter & Delosh, 2006; Chan, McDermott, & Roediger, 2006; Zaromb & Roediger, 2010) should demand much (if not more than restudy) attentional resources and therefore be sensitive to attention division.

How could recent research (Table 1) show detrimental (Dudukovic et al., 2009), null (Gaspelin et al., 2013; Mulligan & Picklesimer, 2016) and beneficial (Kessler et al., 2014) effects of DA during testing on future memory? The discrepancy may lie in experimental details. One notable methodological difference amongst these studies is the duration of each testing trial. Gaspelin et al. (2013) exposed participants to each intervening test trial for up to 19.5 s in Experiment 1 and 8 s in Experiment 2 & 3. Similarly, participants in the two experiments by Mulligan and Picklesimer (2016) restudied or had an intervening test of each word pair for 6 s. In contrast, participants in Dudukovic et al. (2009) had only 3.5 s for each testing trial. A long testing trial duration opens up the opportunity for participants to engage in restudy process (e.g., ruminating on the target word) subsequent to a successful retrieval attempt. Although this extra restudy opportunity would still be susceptible to DA, as one would anticipate from the known effect of DA during a standard (re)encoding, it could also serve to mitigate the detrimental effect of DA during testing on final memory, thereby resulting in the observation of a null effect of DA in these two studies. Indeed, we used the same concurrent task and trial durations as Dudukovic et al. (2009) and the detrimental effect of DA on final test performance was replicated in our Experiment 1. Thus, the contrast between impaired final memory performance with a shorter trial and intact final memory performance with a longer trial as a function of attention division during testing highlights an intriguing hypothesis that awaits future research concerning the important role of re(encoding) processes during testing.

In sum, testing is unlikely a unitary operation but instead may involve memory retrieval, re-encoding as an inherent process of retrieval for future memory, as well as extra restudy if the trial duration is long enough. Although the retrieval itself is largely protected from distraction, the re-encoding part of testing is not resilient to attention division. In other words, focused attention is equally important to both restudy and testing for later memorability, a conclusion that is consistent with other existing theoretical accounts of testing effect.

REFERENCES

- Baddeley, A., Lewis, V. J., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, *113*, 518-540.
- Carpenter, S. K. (2004). A multiple cue hypothesis for the testing effect. *Dissertation Abstracts International, Section B: The Sciences and Engineering*, *66*(1), 584.
- 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*(6), 1563–1569.
- 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*(6), 1547–1552.
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, *19*(5), 619–636.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268-276.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*(4), 553–571.
- Craik, F. M. (2001). Effects of dividing attention on encoding and retrieval processes. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 55-68). Washington, DC, US: American Psychological Association.
- Craik, F. M., 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*(2), 159-180.
- Delaney, P. F., Verkoijen, P. L., & Spigel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory (Vol 53)* (pp. 63-147). San Diego, CA, US: Elsevier Academic Press.
- Dudukovic, N. M., DuBrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory & Cognition*, *37*(7), 953–961.
- Dudukovic, N. (2008). Attending to our memories: The role of attention during declarative memory encoding and retrieval. *Dissertation Abstracts International*, *68*, 6349.
- 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*(2), 155-176.
- Gaspelin, N., Ruthruff, E., & Pashler, H. (2013). Divided attention: An undesirable difficulty in memory retention. *Memory & Cognition*, *41*(7), 978-988.
- Guez, J., & Naveh-Benjamin, M. (2006). Divided attention at encoding and retrieval for once- and thrice-presented items: A micro-level analysis of attentional costs. *European Journal of Cognitive Psychology*, *18*(6), 874–898.

- Hicks, J. L., & Marsh, R. L. (2000). Toward specifying the attentional demands of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1483-1498.
- 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(3): e91309. <https://doi.org/10.1371/journal.pone.0091309>
- Mulligan, N. W., & Picklesimer, M. (2016). Attention and the Testing Effect. *Journal of Experimental Psychology: Learning, Memory, And Cognition*, 42(6), 938-950.
- Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review*, 14(2), 187-193.
- Pu, X., & Tse, C. S. (2014). The influence of intentional versus incidental retrieval practices on the role of recollection in test-enhanced learning. *Cognitive processing*, 15(1), 55-64.
- Pyc, M. A., & Rawson, K. A. (2010). Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science*, 330(6002), 335-335.
- Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181-210.
- Roediger, H. L., & Karpicke, J. D. (2006b). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17(3), 249-255.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463.
- Verkoeijen, P. P., Bouwmeester, S., & Camp, G. (2012). A short-term testing effect in cross-language recognition. *Psychological science*, 23(6), 567-571.
- Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*, 38(8), 995-1008.