

**Retrieval Blocking Supports the Testing Effect with All-of-the-Above Questions on a
Multiple-Choice Test**

by

Bavani Paneerselvam

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
August 1, 2015

Keywords: testing effect, memory retrieval, multiple-choice test

Copyright 2015 by Bavani Paneerselvam

Approved by

Aimee Callender, Assistant Professor of Psychology
Alejandro Lazarte, Associate Professor of Psychology
Ana Franco-Watkins, Associate Professor of Psychology

Abstract

Poorly constructed multiple-choice-questions (MCQ) can have dire consequences on students' exam performance (e.g. negative testing effect). For example, either increasing the answer options (Roediger & Marsh, 2005) or including the correct "none-of-the-above" option (Odegard & Koen, 2007) on a MCQ can lead to poor performance. However the impact of "all-of-the-above" (AOTA) option on the testing effect has not been adequately researched. The present study examined the role of the "all-of-the-above" option on the testing effect. The findings were discussed in light of two types of retrieval hypotheses: retrieval blocking and retrieval fluency. Participants read a set of passages and took an intervening MCQ test for half of the passages and reread half of the passages prior to taking a final MCQ test. Results revealed that previous testing with the correct-AOTA option promoted the testing effect whereas the wrong-AOTA option attenuated the testing effect. This finding is consistent with the retrieval blocking theory. This result will shed light on developing test questions that can help students' retention.

Table of Contents

Abstract.....	ii
List of Tables.....	v
List of Figures.....	vi
Introduction.....	1
Testing Effect.....	1
Theories.....	8
Present Study.....	11
Hypothesis.....	11
Method.....	12
Participants and Design.....	12
Materials.....	13
Procedure.....	13
Results.....	14
Final Test Accuracy.....	14
Conditional Analysis.....	16
Final Test Response Time.....	17
Intervening Test Accuracy.....	18

Discussion	18
References	24
Appendix A	38
Appendix B	39

List of Tables

Table 1.....	31
Table 2.....	32
Table 3.....	33

List of Figures

Figure 1.....	34
Figure 2.....	35
Figure 3.....	36
Figure 4.....	37

Introduction

Test taking reflects not only students' knowledge but also helps in memory retention. For instance, tests can validate students understanding of the material or act as a means to increase retention for a later test. The latter benefit has been termed as the *testing effect* (Roediger & Karpicke, 2006). Furthermore, cognitive psychologists have demonstrated that taking tests repeatedly as a study strategy has advantages that rereading the study material cannot match (Carrier & Pashler, 1992). This robust finding has been replicated with different test formats such as free recall, cued recall, and multiple-choice-questions (MCQ). The benefit associated with the testing effect is the least in the MCQ format compared to other test formats (Foos & Fisher, 1988; Kang, McDermott, & Roediger, 2007). As a result, research on MCQ has been neglected for the most part. However, considering the high prevalence of MCQ testing in educational intuitions, more research on MCQ is found to be emerging. Recent research suggests that MCQ test can in fact produce a bigger testing effect than other test formats depending on the way the questions are constructed (Little, Bjork, Bjork, & Angello, 2012). Furthermore, manipulating lures on an intervening MCQ test can influence the testing effect either positively or negatively (e.g., Jang, Pashler, Huber, 2014; Roediger & Marsh, 2005). Given the diverse effects of testing with MCQ, it is important to determine the different types of MCQ that will benefit retention. The present study examined the effect of an inclusive answer option "all-of-the-above" (AOTA) on the testing effect.

1. Testing Effect

The testing effect paradigm compares a Study-Study-Test condition against a Study-Test-Test condition. The final test can be administered either immediately or after a delay (e.g., 48

hours or 1 week). Roediger and Karpicke (2006) demonstrated the benefits of this testing method among college students. Their study consisted of two groups of college students; the study group and the test group. The study group followed a read-distractor-read sequence and the test group followed a read-distractor-test sequence. After a delay of five minutes, two days, or one week, both the groups were given a final recall test. Counterintuitively, the findings revealed that after a delay of 2 days and 1 week, the test group outperformed the read group on the final test whereas the study group performed better on the final test than the test group after a delay of 5 minutes. Also, it should be noted that in real educational setting students are not tested immediately after a lecture, suggesting that repeated testing may be important for classroom applications. This phenomenon where the repeated testing has positive educational benefit is called testing effect (McDaniel & Masson, 1985; Roediger & Karpicke, 2006; Wheeler & Roediger, 1992).

The advantages of testing are apparent across many domains. Students in a brain and behavior course did better on the final exam when they chose repeated testing as their method of studying compared to restudying the material (McDaniel, Anderson, Derbish, & Morrisette, 2007). Carpenter, Pashler and Cepeda (2007) found that in a visuospatial-geographical test, participants in the testing group were better able to recall the missing information from a map given during the study phase than participants in the regular study condition.

Testing has also been shown to promote transfer of knowledge to new questions within the same knowledge domain. For example, Chan, McDermott and Roediger (2006) tested students on a question (e.g., “Where do toucans sleep at night?”; answer: “In tree holes”) in the initial test phase. On the final test, students were tested with a different set of questions but within the same context (e.g., “What other bird species is the toucan related to?”; answer:

“Woodpeckers”). Once again, final performance on the delayed test was better in the testing condition than the reading condition.

The testing effect improves memory for people of all ages. Interestingly a recent study shows test-taking improves memory in older people. Meyer and Logan (2013) conducted a memory study with an equal number of younger adults and older adults; ranging in age from 55 to 65. The testing group was tested with initial multiple choice questions after reading non-fiction passages. Following the test, they were given feedback on their total score and finally they were tested using short answer questions. Equal testing effect was observed in both the groups.

As observed in lab findings, testing in classrooms produced a similar pattern of the testing effect (McDaniel et al., 2007). For example, this effect is found in middle school (McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011), high school (McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014), university (Leeming, 2002), medical education (Larsen, Butler, & Roediger, 2009), and web-based classroom (McDaniel, Wildman, & Anderson, 2012). Furthermore, this finding is established in different testing formats (Glass & Sinha, 2013).

Concerning the test format, consistency in the formats and questions between the intervening and the final test results in the most robust testing effect (Duchastel & Nungester, 1982). However, some research has shown that testing effect is also present when the final test format differs from the initial test format (McDermott, et al., 2014). For example, Kang et al. (2007) initially tested students with short answer and MCQ, and the final test was given in a MCQ format. So the intervening test in the short answer format resulted in higher retention than

the MCQ format. However, when the final test was given in a short answer format, final test retention was comparable for both the MCQ and short answer intervening tests.

1.1 Multiple Choice Test

Although studies on testing with MCQ were already known in the early 20th century (Spitzer, 1939), only recently the MCQ format has gained considerable attention (Little & Bjork, 2010, 2011). In a typical MCQ testing effect paradigm, the participants read a set of prose passages and take an intervening MCQ test which consist of questions from both the passages that were read and non-read passages. The non-read passages serve as a control condition. Finally they are tested on both the types of passages. The final test format consisted of from free recall, short answers, and MCQ. The general finding shows that performance on an intervening MCQ affects the performance on the final test final differently regardless of the test format (Butler, Marsh, Goode, & Roediger, 2006). For example, the types of options on a MCQ test can either result in a positive testing effect or a negative testing effect (e.g., Odegard & Koen, 2007; Roediger & Marsh, 2005). Thus, these various combinations of the test format coupled with the different options have produced complex results. This complexity is illustrated by manipulation of the lures on the intervening test.

1.1.1 Cross-format MCQ

Presence of additional lures generally does not facilitate retention for prose passages (Butler et al., 2006) but facilitates retention for non-related word stimuli (Whitten & Leonard, 1980). For prose passages, the fewer the number of lures on the intervening test, the greater is the testing effect (Roediger & Marsh, 2005). Also, students use the lures from the intervening tests

to answer the final cued-recall test (Fazio, Agarwal, Marsh, & Roediger, 2010; Roediger & Marsh, 2005).

For example, Roediger and Marsh (2005) had participants read prose passages and take an intervening MCQ test where the number of lures were manipulated. After that, participants took a cued-recall test. Their results indicated that higher the number of lures on the intervening test, the worse is their final accuracy which, thereby negating the testing effect. Furthermore, participants used the lures from the intervening test as answers on the final test.

Odegard and Koen (2007, Exp.1) extended upon Roediger and Marsh's study and examined the testing effect of inclusive options, "none-of-the-above" (NOTA) on the intervening MCQ test. In experiment 1, they had three different types of questions on the intervening test. The first type was the standard format (4 options: 1 target; 3 lures). The other two were in the inclusive format with the correct none-of-the-above (5 options: 4 lures; none-of-the-above) and the wrong none-of-the-above (5 options: 1 target; 3 lures; none-of-the-above). The correct-NOTA exposed students to only the lures. However, the wrong-NOTA exposed students to 3 lures and the target. The final cued-recall test had the exact questions from the intervening MCQ test. Results indicate that the initial testing with the correct-NOTA negated the testing effect and students used the lures from the intervening test as answers on the final test.

Using another inclusive answer option "all-of-the-above" (AOTA), Bishara and Lanzo (2014, Exp.1) examined the testing effect on a final cued-recall test when the intervening test included AOTA questions. In a series of experiments, they had participants read prose passages and take an intervening test under four different conditions: no test (control questions), standard test (4 options: 1 target; 3 lures), correct all-of-the-above (5 options: 4 targets; all-of-the-above), and wrong all-of-the-above (5 options: 1 target; 3 lures; all-of-the-above) before taking a final

cued-recall test with the same questions from the intervening test. Results show that prior testing with the correct-AOTA enhanced the testing effect. Also, lure intrusion from the intervening test was the lowest for the correct-AOTA questions on the final test because correct-AOTA does not expose students to lures compared to the standard questions or the wrong-AOTA questions.

Taken together these studies reach a similar conclusion that the presence of lures on the intervening test negates the testing effect whereas the absence of lures enhances the testing effect in cross-format testing.

1.1.2 Same format MCQ

Considering the consistency in formats between the intervening and the final test in classroom settings, is it important to know whether manipulation of the lures on the intervening MCQ test benefit or cost retention on the final MCQ test.

Odegard and Koen (2007, Exp. 2) manipulated the answer options on the intervening MCQ test and the final MCQ test in their investigation of none-of-the-above questions. Their intervening test manipulation is the same as in experiment one, [standard format (4 options: 1 target; 3 lures), correct none-of-the-above (5 options: 4 lures; none-of-the-above) and the wrong none-of-the-above (5 options: 1 target; 3 lures; none-of-the-above)]. The control questions were questions taken from passages that were not read during the study phase but tested on the intervening test. However, the final test was in the standard MCQ format (4 options: 1 target; 3 lures) with the same questions from the intervening test. They found that questions for which none-of-the-above was the correct option (lures present), these items negated the testing effect. There was no marked difference between the final accuracy of the initially tested wrong none-of-the-above questions and the control questions.

Bishara and Lanzo (2014, Exp. 2) manipulated the answer options on the intervening test and the final multiple-choice test with AOTA questions. The intervening test manipulation was the same as in experiment one, [no test (control condition), standard test (4 options: 1 target; 3 lures), correct-AOTA (5 options: 4 lures, all-of-the-above) and wrong-AOTA (5 options: 1 target, 3 lures, all-of-the-above)]. The control questions were taken from the read passages but not tested on the intervening test. The final multiple-choice test consisted of 5 option questions: correct-AOTA (5 options: 4 lures; all-of-the-above) and wrong-AOTA (5 options: 1 target; 3 lures; all-of-the-above). They found that for the wrong-AOTA questions on the final test, previously tested with the correct-AOTA on the intervening test lead to higher retention than the no-test (control questions), standard format, and the wrong-AOTA questions on the intervening test. For the correct-AOTA questions on the final test, although previously tested with the correct-AOTA questions on the intervening test lead to higher retention on the final test, it was not significantly different than the no-test (control) questions.

These studies show that the presence of the additional lures negates the testing effect, similar to the cross-format testing. Although these findings confirm the link between the lures and the testing effect, the methodological flaws in these studies should be taken into account. It should be noted that none of these studies compared the testing group with a proper restudy control, and studies that did, failed to show the testing effect (Carpenter & DeLosh, 2006). The second limitation is the absence of a delay condition. The testing effect is more evident after a delay, so testing with additional lures might have different effects when the final test is given after a few days. Nevertheless, these studies imply that additional lures on the intervening MCQ test affect the testing effect regardless of the final test format.

2. Theories

Theories of the Testing Effect:

Amount of Processing Hypothesis/ Additional Exposure Theory

The most elementary theory of the testing effect is the *additional exposure theory*. Thompson, Wenger, and Bartling (1978) erroneously argued that prior exposure to the tested material in both the study and test conditions promotes the testing effect. Testing effect was evident even in studies that provided equal exposure to both the restudy and the testing group. Therefore, additional exposure theory failed to explain the underlying cognitive mechanism and was dismissed.

Retrieval theory

This theory proposes that the act of retrieval increases retention (Dempster, 1996). The majority of the testing effect studies have examined the retrieval theory using cued-recall test (e.g., Carrier & Pashler, 1992), and very few very studies have used multiple-choice test (Bishara & Lanzo, 2014; Jang et al., 2014). Although, broadly retrieval theory has been implicated with multiple-choice test, the existing studies are lacking in terms of specifying a particular theory for the divergent finding (i.e. positive testing effect, negative testing effect) when tested with different types of MCQ. Therefore, under the umbrella of the retrieval theory, several different theories are outlined to account for the testing effect with AOTA questions.

a) *Retrieval blocking hypothesis*

This theory posits that retrieval and exposure to the lures can block the retrieval of the targets on the final test and, thereby blocking the accessibility of the original information (Raaijmakers & Shiffrin, 1981; see also Roediger & Neely, 1982). Although this hypothesis

stems from part-list cuing paradigm (Slamecka, 1968; 1969), the support for this theory is also observed in paired-associate paradigm (Mueller & Watkins, 1977).

Besides the evidence of this theory in different paradigm, the evidence linking retrieval blocking and the misinformation effect in eyewitness memory is compelling. MCQ that has the lures as options is analogous to the misinformation paradigm (Roediger & Marsh, 2005). In a classic misinformation study conducted by Loftus, Miller, and Burns (1978), participants viewed a slide of a car stopping at “yield” sign. Later the group was divided into a control and misinformed group. The control group reread a description of the original information. However, the misinformed group read novel description about the car stopping at a “stop” sign. Similarly, the lures viewed on the intervening MCQ can impair retrieval of the original information.

For example, Schreiber and Sergent (1998) asked the participants to view slides of a crime and exposed them to a misinformed description. Later the experimental group was given an intervening recognition test that had lures. Finally, they were given another recognition test measuring the original retention for the slide. In contrast to the control group, the experimental group performed poorly on the final test. In essence, the negative testing effect was observed when retrieval practice was coupled with lures. Their finding is in accordance to the retrieval blocking hypothesis where the retrieval of the lures blocked the retrieval of the targets.

Also supporting the retrieval blocking hypothesis and the misinformation effect is the striking finding that comes from educationally relevant material. For example, Odegard and Koen (2007) examined the role of “none-of-the-above” (NOTA) option on the testing effect. They found that having previously tested with the correct-NOTA (lures present) negated the testing effect and the lures were given as answers on the final test. When the lures were retrieved on the intervening correct-NOTA questions, the retrieval of the targets was blocked. They further

added that this pattern could be reversed through exposure to only the targets on an intervening test. In view of their prediction, testing with the correct-AOTA (lures absent) on the intervening test should block the retrieval of the lures on the final test

b) *Retrieval fluency hypothesis*

The premise of this theory is that if an answer comes to the mind easily, then that answer will be chosen (Baddeley, 1982; Benjamin, Bjork, & Schwartz, 1998; Jacoby & Dallas, 1981).

One explanation for the underlying mechanism involved is exposure and familiarity. Several studies have pointed out that participants feel a statement is true because of its prior exposure (e.g., Zaragoza & Lane, 1994; Zaragoza & Mitchell, 1996). For example, participants in Kelley and Lindsay (1993) were tested on general knowledge questions. Prior to the memory test, they were exposed to the targets and lures. In a series of experiments, they found that the previously exposed lures were chosen confidently regardless of pre-exposure warning (“some of the answers are wrong”). They argued that the exposure to the lures lead to retrieval fluency.

Other evidence comes from studies of eyewitness memory. Zaragoza and Mitchell (1996) manipulated the number of exposure to the suggestive information in the questions and later tested the participants’ memory of the original event. Similar to Kelley and Lindsay, they found that prior single and multiple exposures to the lures caused the participants to claim with confidence that they remember the misleading events, and this pattern persisted even after a delay of one week.

Considering the parallel between the misinformation effect and the MCQ, it is possible that the retrieval fluency is used while answering MCQ. The above studies show that prior exposure to the misinformation increased the fluency for the misinformation; similarly then, prior exposure to the correct information should increase fluency for the correct information. The

correct-AOTA, which has only the targets on the intervening test, should increase the saliency of the targets on the final test that has a mixture of lures and targets

For example, the correct-AOTA questions on the intervening test consist of only the targets. Prior exposure to the targets should increase the targets saliency in a pool of lures on the final test; thereby the target is retrieved fluently. In contrast, the wrong-AOTA questions have a mixture of lures and targets on both the intervening and the final test. The saliency of both the targets and lures on the final test should impede the fluent retrieval of the targets.

3. Present Study

In sum, the previous studies showed that the lures on the intervening test can affect the testing effect. On one hand, inclusion of the lures (i.e. NOTA) on the intervening test negates the testing effect on the subsequent test (Odegard & Koen, 2007). On the other hand, absence of lures (i.e. AOTA) on the intervening test enhances the testing effect (Bishara & Lanzo, 2014). Importantly, the cognitive mechanism underlying the testing benefit from the correct-AOTA remains unclear. The present study extended the previous study on the testing effect with AOTA questions with regards to two different retrieval theories: retrieval blocking and retrieval fluency, combining a refined methodology with a restudy control group, a delay (48 hours) condition, an intervening MCQ test in an inclusive 5 options format (correct-AOTA: 4 lures, all-of-the-above) and (wrong-AOTA: 1 target, 3 lures, all-of-the-above), and a final MCQ test in a standard format (4 options: 1 target; 4 lures).

3.1. *Hypothesis*

The following hypotheses were generated after a review of the above-cited literature:

Hypothesis 1: Of primary interest, I hypothesized that there will be a testing effect for the critical questions when previously tested with the correct-AOTA on the intervening test. That is, the final test accuracy for the critical questions should be higher than the control questions when previously tested with the correct-AOTA questions. The support for this hypothesis is retrieval blocking. The absence of lures on the correct-AOTA questions should block the retrieval of the lures.

Hypothesis 2: I hypothesized that there will be no testing effect for the critical questions when previously tested with the wrong-AOTA on the intervening tests. That is, the final test accuracy for the critical questions should be lower than or equivalent the control questions when previously tested with the wrong AOTA questions. The support for this hypothesis is retrieval blocking. The presence of lures on the wrong-AOTA questions should block the retrieval of the targets.

Hypothesis 3: Additionally, I hypothesized that the response time for the critical correct-AOTA questions will be shorter than the critical wrong-AOTA questions. That is, prior exposure to the correct-AOTA on the intervening test should increase familiarity to the targets, thereby increasing the retrieval fluency for the targets on the final test. The support for this hypothesis is retrieval fluency.

4.1. Method

4.1.1 Participants & Design

101 undergraduates from Auburn University completed the experiment as part of course requirement. 3 participants were excluded from data analysis due to missing data. The experiment had a 2 (timing: immediate, delay) x 2 (item type: control, tested) x 2 (AOTA type:

correct, wrong) mixed ANOVA design. Timing was manipulated between subjects. The full design is presented in Figure 1.

4.1.2 Materials

Study item consist of 16 non-fiction passages (8 critical; 8 control) adapted from the Graduate Record Examination (GRE) and Test of English as a Foreign Language (TOEFL) taken from Roediger and Marsh (2005).

The intervening multiple-choice test assessed participants' knowledge of the 8 critical passages read. The intervening test followed the AOTA test format (5 options) and included 4 questions from each passage (32 questions total). From the 4 questions presented for each passage, 2 questions had AOTA as the correct answer and the other 2 questions had AOTA as the wrong answer. So, in total, 16 critical questions had AOTA as the correct answer and other 16 critical questions had AOTA as the wrong answer.

The final multiple-choice test followed the standard format (4 options) with the exact critical questions from the intervening test plus the control questions but without the AOTA option. The final test consists of 64 questions: 16 critical questions corresponded to the AOTA correct and wrong intervening questions, 16 questions corresponded to the AOTA incorrect intervening questions, and 32 questions were control questions from the passages that were read but not tested during the intervening test. Both the passages and the tests will be presented on *E-prime*.

4.1.3 Procedure

Participants were recruited from the Auburn University's SONA subject pool. On the day of their scheduled time, informed or parental consent and demographics questionnaire (age, gender, and race) were obtained from all participants prior to beginning the study.

The experiment was broken into two different counterbalancing conditions. Participants were randomly assigned to one of the conditions.

Encoding phase, all participants studied 16 passages in total. 8 of the passages served as critical passages (tested passages) and 8 as control passages. The passages were counterbalanced so that the 8 passages that are critical passages for one condition were the control passages for the other condition and vice versa. They were given 90 seconds to complete each passage.

Overall, this phase took 24 minutes.

Intervening phase, participants in both counterbalancing conditions were tested on the 8 critical passages and restudied the 8 control passages from their respective encoding phase. For both counterbalancing conditions, participants alternated between taking a test from a critical passage and rereading a control passage. Participants were tested on 32 questions in total, with 4 questions from each of the 8 critical passages. In order to equate the exposure time, they were given 90 seconds to complete each passage and test. Overall, this phase took 24 minutes.

Final test phase, all the participants took the same final test (questions were in the same order for all participants) with a total of 64 questions. They were tested on 32 questions from the critical passages and 32 from the control passages read during the encoding phase. Participants in the immediate condition took the final test after a delay of 5 minutes whereas participants in the delay condition took the final test after a delay of 48 hours.

4.2 Results

Final test accuracy

The main aim of the study was to examine whether initial testing with different AOTA questions increases the testing effect. Table 1 reports proportions of accuracy on the final test as a function of item type, AOTA type, and timing.

A 2 (item type: critical, control) x 2 (AOTA type: correct- AOTA, wrong- AOTA) x 2 timing (immediate, delay) mixed ANOVA reveals main effects for item type, $F(1, 96) = 34.817$, $MSE = .54$, $p < 0.001$, $\eta_p^2 = .23$, AOTA type, $F(1, 96) = 115.301$, $MSE = 1.53$, $p < 0.001$, $\eta_p^2 = .55$, and timing, $F(1, 96) = 10.94$, $MSE = .10$, $p = .001$, $\eta_p^2 = .10$. Consistent with the testing effect, participants scored higher for the previously tested questions, critical questions ($M = .64$) than the restudy condition, control questions ($M = .57$). Importantly, participants scored higher for critical correct-AOTA ($M = .67$) than critical wrong-AOTA ($M = .54$). Unsurprisingly, they scored higher in the immediate condition ($M = .63$) than the delay condition ($M = .57$). This was qualified by item type x AOTA interaction, $F(1, 96) = 8.23$, $MSE = .19$, $p = 0.005$, $\eta_p^2 = .08$.

Pairwise comparison revealed a significant difference between the critical ($M = .72$) and the control questions ($M = .61$) for correct-AOTA. In contrast, no significant difference between the critical ($M = .56$) and the control questions ($M = .52$) for the wrong-AOTA questions. The absence of method x AOTA x timing interaction indicates that this pattern persisted even after a delay, see Figure 2.

Thus the result shows that initial testing with the AOTA questions increased the final test accuracy. Specifically initial testing with the correct-AOTA enhanced the testing effect. Lack of similar result for the control questions confirms that participants engaged in retrieval strategy while answering AOTA questions.

Conditional Analysis

Conditional analysis examined the relationship between the initial test performances and the final test performance. Specifically, how the absence of incorrect lures (correct-AOTA) or presence of incorrect lures (wrong-AOTA) on the intervening test coupled with initial retrieval affected the final test accuracy. For the correct-AOTA, the absence of lures should block retrieval of lures on the final test, whereas for the wrong-AOTA, the presence of lures should block the retrieval of target. If this is so, then blocking should account for the underlying mechanism.

How does retrieving either the target or lure on the intervening test affect the final test accuracy? Table 2 shows the proportion of accuracy on the final test as a function of initial retrieval on the intervening test and AOTA type. A 2 (initial retrieval: target, lure) x 2 (AOTA type: correct-AOTA, wrong-AOTA) x 2 (timing: immediate, delay) mixed ANOVA reveals main effects for initial selection, $F(1,96) = 223.87$, $MSE = 10.36$, $p < 0.001$, $\eta_p^2 = 0.7$, AOTA type, $F(1,96) = 93.105$, $MSE = 3.46$, $p = 0.01$, $\eta_p^2 = 0.49$, and timing, $F(1,96) = 4210.05$, $MSE = 0.23$, $p = 0.015$, $\eta_p^2 = .06$, and initial retrieval x AOTA type interaction $F(1,96) = 88.92$, $MSE = 3.15$, $p = .001$, $\eta_p^2 = .48$. If the target was retrieved on the initial test, the probability of retrieving the target on the final test was equal for both the correct-AOTA ($M = .79$) and the wrong-AOTA questions ($M = .79$). However, when lures are retrieved on the intervening test, the probability of retrieving the target was higher for the correct-AOTA questions ($M = .65$) than the wrong-AOTA questions ($M = .28$). See figure 3.

In summary, for both the correct-AOTA questions and wrong-AOTA questions, retrieval of the target on the intervening test blocked the retrieval of lures. However, when the lures were

retrieved on the intervening test, it had opposing effects for the AOTA questions. For the correct-AOTA questions, more targets were retrieved than for the wrong-AOTA questions.

Final test response time

Response time was analyzed to examine whether retrieval fluency accounts for the higher accuracy on the correct-AOTA questions. Data from 3 participants were excluded because they were outliers. Reaction time above 3 standard deviations was used as the exclusion criteria to trim the outliers. Table 2 reports mean response time on the final test as a function of item type, AOTA type, and timing.

A 2 (item type: critical, control) x 2 (AOTA type: correct- AOTA, wrong- AOTA) x 2 (timing: immediate vs. delay) mixed ANOVA reveals main effects for the item type, $F(1, 93) = 57.55$, $MSE = 8.61$, $p < 0.001$, $\eta_p^2 = .38$, AOTA type, $F(1,93) = 140.56$, $MSE = 5.10$, $p < 0.001$, $\eta_p^2 = .60$, and timing, $F(1,93) = 10.70$, $MSE = 5.83$, $p = .002$, $\eta_p^2 = .10$. As expected, participants spent less time on the critical questions ($M = 2.22$) than the control questions ($M = 1.92$) and immediately ($M = 1.95$) than after a delay ($M = 2.20$). Unexpectedly, participants spent longer on the correct-AOTA ($M = 2.19$) than wrong-AOTA ($M = 1.96$). This was qualified by item type x AOTA type interaction, $F(1, 93) = 8.35$, $MSE = 0.32$, $p = 0.005$, $\eta_p^2 = .08$.

Pairwise comparison revealed a significant difference between the correct-AOTA ($M = 2.31$) and the wrong-AOTA ($M = 2.14$) for the control questions. Similarly, there was a significant difference between correct-AOTA ($M = 2.07$) and wrong-AOTA ($M = 1.78$) for the critical questions. See Figure 4.

Although overall participants retrieved the answers quickly for the critical questions, the answer for the correct-AOTA questions was retrieved slower than wrong-AOTA questions,

against the prediction. Thus, the retrieval fluency theory does not support the enhanced testing effect with the correct-AOTA. However, retrieval fluency is operating for the wrong-AOTA questions, which is evidenced by a larger difference between the question types for the critical questions and the control questions.

Intervening test Accuracy

A 2 (AOTA type: correct, wrong) x 2 (Timing: immediate, delay) mixed ANOVA resulted in a main effect for the AOTA, $F(1, 96) = 14.602$, $MSE = .33$, $p < .001$, $\eta_p^2 = .13$. Participants scored significantly higher for the correct AOTA ($M = .59$) than the wrong AOTA ($M = .51$). AOTA x timing interaction did not reach significance, $F(1, 96) = 0.397$, $MSE = .01$, $p = .53$, $\eta_p^2 = 0.004$.

Intervening test response time

A 2 (AOTA type: correct, wrong) x 2 (Timing: immediate, delay) mixed ANOVA found no significant main effect of AOTA type, $F(1, 96) = .70$, $MSE = .06$, $p = .43$, $\eta_p^2 = .01$ or AOTA x timing interaction, $F(1, 96) = .02$, $MSE = .00$, $p = .88$, $\eta_p^2 = .00$. Overall, participants spent equivalent time to answer both the correct-AOTA and wrong- AOTA questions.

4.3 Discussion

The results show that the testing effect is enhanced when previously tested with the correct-AOTA questions, replicating previous research (Bishara & Lanzo, 2014) and extending this benefit across time with a pure restudy control design. Additionally, this study provides support that the retrieval blocking theory contributes to the overall testing effect for the AOTA questions. Prior testing with the correct-AOTA questions resulted in higher retention for that question on the final test compared to control questions, indicating a positive testing effect.

However, when previously tested with the wrong-AOTA questions, retention is comparable to the control questions, indicating no testing effect. Typically the testing effect is present when the final test is given after a delay of more than 24 hours (Roediger & Karpicke, 2006). However, in this study the testing effect is observed both immediately and after a delay of 48 hours, consistent with few testing effect studies on MCQ test (e.g., Little & Bjork, 2012).

Presence of lures on an intervening test has been shown to affect the final test accuracy (Butler et al., 2006). Conditional analysis was conducted to determine how the presence of lures (wrong-AOTA questions) or absence of lures (correct-AOTA questions) affected the intervening test accuracy, which in turn affected the final test accuracy. When targets were retrieved on the intervening test, the targets were consistently retrieved on the final test, and this was observed for both the correct-AOTA and the wrong-AOTA questions. This is unsurprising because students hold on to the original answer as a byproduct of commitment (Benjamin, Cavell, & Shallenberger, 1984). Interestingly, however, when lures were retrieved on the intervening test, the commitment effect disappears for the correct-AOTA questions. That is, for the correct-AOTA questions, answers were changed from the incorrect lures on the intervening test to targets on the final test. To illustrate, for the correct-AOTA questions: “Where do vampire bats make incision?”, the options (toes, shoulder, ears, neck, all-of-the-above) are all correct facts. Therefore, when the students’ retrieved the correct answer (all-of-the-above) or the incorrect answer (toes, shoulder, ears, neck), they have nevertheless retrieved only the correct facts. As a result, the absence of lures on the correct-AOTA question blocked the retrieval of the lures on the final test. This is evidenced by the results showing that regardless of the answer chosen on the initial test, there is no decrement on the final test accuracy. In contrast, the wrong-AOTA question consists of lures and a target. For example, the wrong-AOTA question: Bats are

classified as part of which of the following mammalian order?, the options (chiroptera, insectivore, marsupialia, primata, all of the above) has chiroptera as its target and insectivore, marsupialia, and primata as its incorrect lures. In addition to retrieving the incorrect lures on the intervening test, the presence of other incorrect lures on the intervening test blocked the retrieval of the target on the final test and lead to decrement of accuracy on the final test (Odegard & Koen, 2007). Thus, this result pinpoints that the retrieval blocking contributed for the enhanced testing effect for the correct-AOTA questions and the null testing effect for the wrong-AOTA questions. This finding confirms the hypothesis one, that the final test accuracy for the critical questions should be higher than the control questions when previously tested with the correct-AOTA questions. Also, this finding confirms the hypothesis 2, that the final test accuracy for the critical questions should be lower than or equivalent to the control questions when previously tested with the wrong-AOTA questions.

On the intervening test, accuracy was higher for the correct-AOTA than the wrong-AOTA questions. One might argue that with the correct-AOTA questions, students can guess *all-of-the-above* as the logical answer once they are sure with at least two targets. However, with the wrong-AOTA questions, knowing two lures discounts all-of-the-above as the answer and forces them to search for the target. If this is so, then the time taken to search for the target should be longer. However, data from the reaction time shows that students spent the same amount of time retrieving answers for both the correct-AOTA questions and the wrong-AOTA questions. This also shows that participants are reading through all of the options and not just choosing all-of-the-above when 2 targets are recognized. Therefore, retrieval fluency on the intervening test did not account for the testing effect on the final test. However the level of processing might have affected the final test accuracy. Butler et al. (2006) provided evidence that students engage in

different level of processing on a MCQ test with lures. Although item-specific processing is implicated with recognition test in general (Eisntein & Hunt, 1980), inclusive answer options on a MCQ test might involve relational processing. Questions with inclusive options by its nature require a compare and contrast method among the options. This method is harmful for the wrong-AOTA questions because in the process of evaluating the options, a strong connection is established between the targets and the lures, that later interferes with retrieval on the final test. That is, the connection between the options on the MCQ test increased performance for the correct-AOTA questions but performance decreased for the wrong-AOTA questions. Therefore, this shows that students engaged in a similar process for both the correct-AOTA and the wrong-AOTA questions on the intervening test. However, the presence of lures on the intervening questions is what affected the final test accuracy.

Typically, the testing effect with MCQ test is prominent only when the control is a non-tested question (Rowland, 2014). However, this study showed the testing effect with a pure restudy control. To rule out that more exposure on the test questions lead to the testing effect, we equated the exposure time for both the restudy passages and test questions. Despite that, students did better for the tested questions than the control questions, especially for the correct- AOTA questions. This finding further supports the existing studies showing that the additional exposure theory does not account for the testing effect (Roediger & Karpicke, 2006).

The enhanced testing effect with the correct-AOTA questions cannot be attributed to the retrieval fluency, as measured by the time taken to answer the correct-AOTA questions. In fact, students spent a longer time retrieving the answers for the correct-AOTA questions than for the wrong-AOTA questions. However the null testing effect for the wrong-AOTA can be attributed to the retrieval fluency. One possible explanation is that since the final test had novel lures for

the correct-AOTA questions, students' engaged in an elaborative processing to discriminate the target from the lures. However, the same lures on the wrong-AOTA questions lead to fluent retrieval of the answers on the final test (Jacoby, Shimizu, Daniels, & Rhodes, 2005), bypassing the elaborative processing. Although some studies have shown retesting with the old vs novel lures produce similar testing effect (Jang et al., 2014; Odegard & Koen, 2007), other studies have shown the opposite (Rees, 1986). This finding contradicts hypothesis 3, that the prior exposure to the correct-AOTA on the intervening test should increase familiarity to the targets, thereby increasing the retrieval fluency for the targets on the final test.

The co-existence of recollection and familiarity processes with the recognition test (Yonelinas, 2002), might have enabled different processes for the different types of AOTA questions. For the wrong-AOTA questions, familiarity processes were operating with old lures and participants using recollection for the correct-AOTA questions with novel lures. Although previous study showed that similar testing effects are observed regardless of the lure types, but the question whether or not the retrieval blocking theory and the retrieval fluency theory can account for the testing effect using different lures still remains a challenge. It is thus important to standardize the lures in future studies. Another limitation of this study is that no test in the standard format was used in both the intervening and the final test. This makes it difficult to know if testing in the inclusive options is more beneficial than the standard format. Also, it is difficult to pinpoint if students engage in a different strategy while answering questions on a standard format compared to the inclusive options format. Future studies should include standard formats in addition the inclusive formats.

This is the first study to provide theoretical support for the enhanced testing effect with the correct-AOTA questions on the MCQ test. A plethora of studies have focused on finding the

optimal test format to reap the highest retention. In comparison, very few studies have focused on same test format to discover the best manipulation for higher retention. Alternating the formats within the broader MCQ format can not only enhance the testing effect with the AOTA questions, it can also stimulate challenging cognitive processes. This study has important pedagogical implications for popularly used MCQ testing formats in the classroom, highlighting that AOTA questions facilitate retention of the material when tests are given on a heterogeneous MCQ formats.

References

- Baddeley, A. D. (1982). Domains of recollection. *Psychological Review*, *89*, 708–729.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68.
- Benjamin, L. T., Cavell, T. A., & Shallenberger, W. R. (1984). Staying with initial answers on objective tests: Is it a myth? *Teaching of Psychology*, *11*, 133–141
- Bishara, A. J., & Lanzo, L. A. (2014). All of the above: When multiple correct response options enhance the testing effect. *Memory*, 1-16.
- Butler, A. C., Marsh, E. J., Goode, M. K., & Roediger, H. L., III (2006). When additional multiple-choice lures aid versus hinder later memory. *Applied Cognitive Psychology*, *20*, 941-956.
- 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*, 268-276.
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*, 760–771.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, *20*, 633-642.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, *135*, 553–571.

- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition: Memory* (pp. 317–344). San Diego, CA: Academic Press.
- Duchastel, P. C., & Nungester, R. J. (1982). Testing effects measures with alternate test forms. *Journal of Educational Research, 75*, 309–313.
- Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588–598.
- Fazio, L. K., Agarwal, P. K., Marsh, E. J., & Roediger, H. L., III. (2010). Memorial consequences of multiple-choice testing on immediate and delayed tests. *Memory & Cognition, 38*, 407–418.
- Foos, P. W., & Fisher, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology, 80*, 179–183.
- Glass, A. L., & Sinha, N. (2013). Multiple-choice questioning is an efficient instructional methodology that may be widely implemented in academic courses to improve exam performance. *Current Directions in Psychological Science, 22*, 471–477.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.
- Jang, Y., Pashler, H., & Huber, D. E. (2014). Manipulations of choice familiarity in multiple-choice testing support a retrieval practice account of the testing effect. *Journal Of Educational Psychology, 106*(2), 435-447.

- Kang, S. H. K., McDermott, K. B., Roediger, H. L., III. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology, 19*, 528–558.
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1–24.
- Larsen, D. P., Butler, A. C., & Roediger, H. L., III (2009). Repeated testing improves long-term retention relative to repeated study: A randomized, controlled trial. *Medical Education, 43*, 1174–1181.
- Leeming F.C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210–212.
- Little, J. L., & Bjork, E. L. (2010). Multiple-choice testing can improve the retention of non-tested related information. In S. Ohisson & R. Catrabone (Eds.) Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 1535-1540). Austin, TX: Cognitive Science Society.
- Little, J. L., & Bjork, E. L. (2011). Pretesting with multiple choice questions facilitates learning. In L. Carlson, C. Hölischer & T. F. Shipley. (Eds.) Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 294-296). Austin, TX: Cognitive Science Society.
- Little, J. L., & Bjork, E. L. (2012). The persisting benefits of using multiple-choice tests as learning events. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society (pp. 683–688). Austin, TX: Cognitive Science Society.
- Little, L. J., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-Choice Tests Exonerated,

- at Least of Some Charges: Fostering Test-Induced Learning and Avoiding Test-Induced Forgetting. *Psychological Science*, 23, 337-344.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19-31
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103, 399 - 414.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513.
- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 371-385.
- McDaniel, M. A., Wildman, K. M., Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, 1, 18-26.
- McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, 20, 3-21.
- Meyer, A.N.D. & Logan, J.M. (2013). Taking the Testing Effect Beyond the College Freshman: Benefits for Lifelong Learning. *Psychology and Aging*, 28, 142-147.
- Mueller, C. W., & Watkins, M. J. (1977). Inhibition from part-set cuing: A cue-overload interpretation. *Journal of Verbal Learning & Verbal Behavior*, 16, 699-709.
- Odegard, T. N., & Koen, J. D. (2007). "None of the above" as a correct and incorrect alternative

- on a multiple-choice test: Implications for the testing effect. *Memory*, *15*, 873-885.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93-134.
- Roediger, H.L. & Neely, J.H. (1982). Retrieval blocks in episodic and semantic memory. *Canadian Journal of Psychology*, *36*, 231-242.
- 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.
- Roediger, H. L. III, & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1154_1159.
- 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
- Schreiber, T. A., & Sergent, S. D. (1998). The role of commitment in producing misinformation effects in eyewitness memory. *Psychonomic Bulletin & Review*, *5*, 443–448.
- Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, *76*, 540-513.
- Slamecka, N. J. (1969). Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology*, *81*, 557-560.
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, *30*, 641–656.
- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 210–221.

- Wheeler, M. A., & Roediger, H. L., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*, 240-245.
- Whitten, W. B., & Leonard, J. M. (1980). Learning from tests: facilitation of delayed recall by initial recognition alternatives. *Journal of Experimental Psychology: Human Learning and Memory, 6* (2), 127-134.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517
- Zaragoza, M. S., & Lane, S. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 934-945.
- Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestions and the creation of false memories. *Psychological Science, 7*, 294-300.

Table 1. Proportion correct on the final test as a function of initial item type, initial AOTA type, and timing (with Standard Deviations in Parentheses).

Item type x AOTA type	Timing	
	Immediate	Delay
Critical correct-AOTA	0.75 (0.14)	0.70 (0.13)
Control correct-AOTA	0.63 (0.13)	0.59 (0.14)
Critical wrong-AOTA	0.59 (0.16)	0.52 (0.15)
Control wrong-AOTA	0.58 (0.14)	0.47 (0.19)

Note. AOTA = all-of-the-above.

Table 2. Probability of accuracy on the final test as a function of initial selection type, initial AOTA type, and timing (with Standard Deviations in Parentheses).

Initial selection type x AOTA type	Timing	
	Immediate	Delay
<i>correct-AOTA</i>		
p(target target)	0.81 (0.16)	0.78 (0.15)
p(lure target)	0.67 (0.22)	0.62 (0.21)
<i>wrong-AOTA</i>		
p(target target)	0.81 (0.18)	0.79 (0.19)
p(lure target)	0.32 (0.24)	0.25 (0.21)

Note. AOTA = all-of-the-above.

Table 3. Mean response time in minutes on the final test as a function of initial item type and initial AOTA type (with Standard Deviations in Parentheses).

Item type x AOTA type	Timing	
	Immediate	Delay
Critical correct-AOTA	1.94 (.35)	2.19 (.42)
Control correct-AOTA	2.21 (.50)	2.41 (.48)
Critical wrong-AOTA	1.60 (.34)	1.95 (.46)
Control wrong-AOTA	2.04 (0.44)	2.23 (.51)

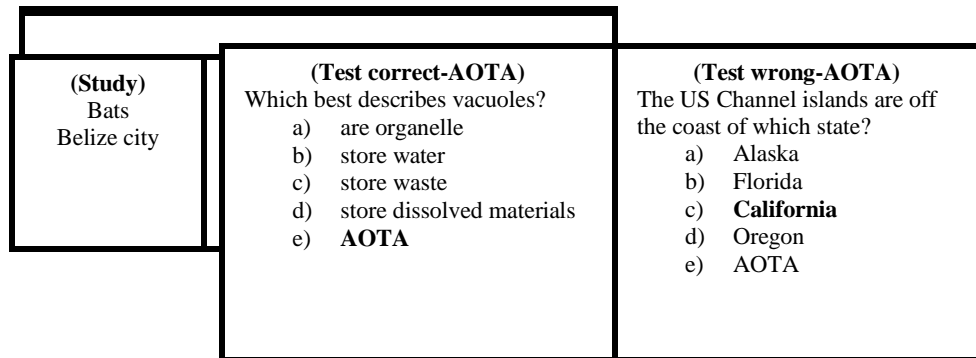
Note. AOTA = all-of-the-above.

Encoding

(Study)
Bats
Belize city
Cells
Chumash



Intervening Phase



Final

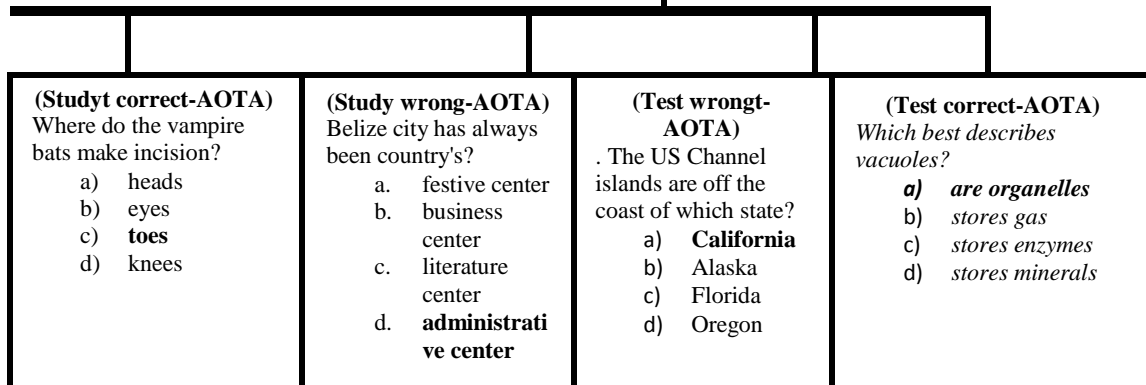


Figure 1. Schematic representation of the experimental design. The answers for each question types are bolded.

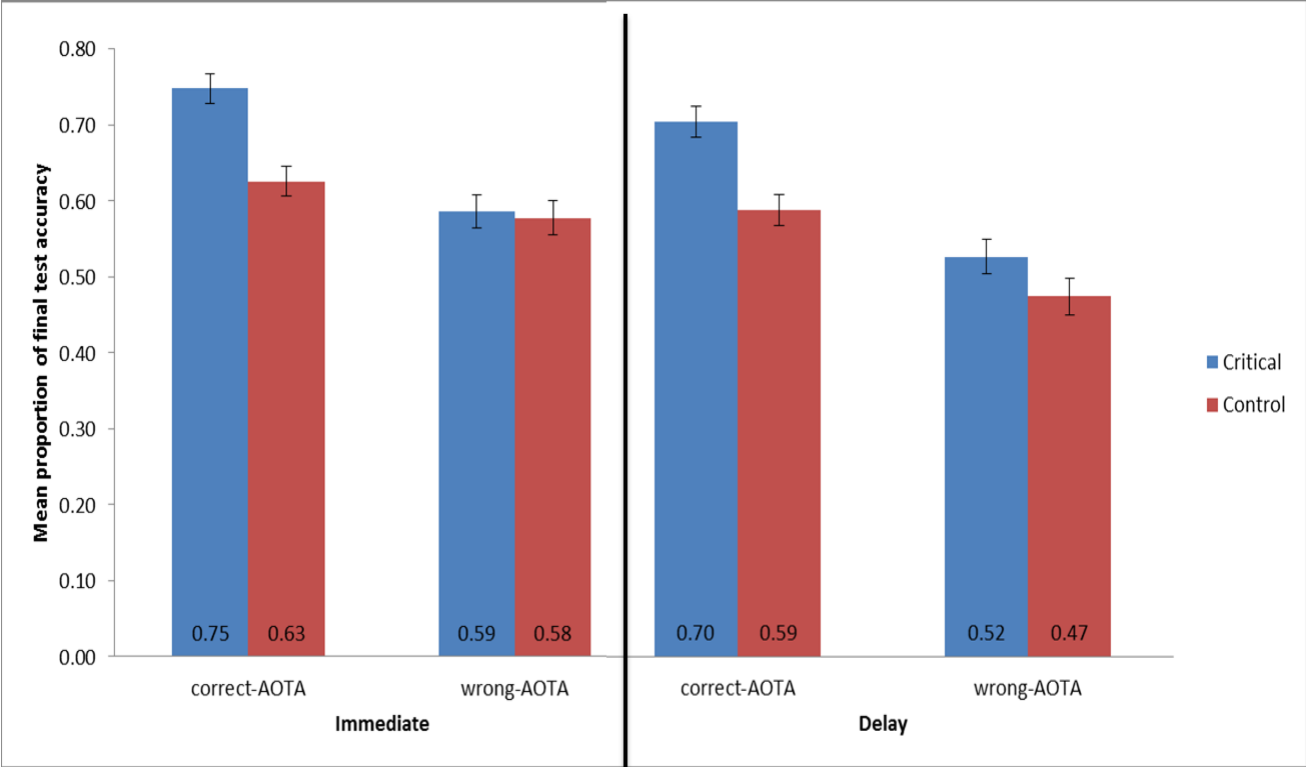


Figure 2. Mean proportion on the final test accuracy as a function of AOTA type, item type, and timing. Bars represent standard error.

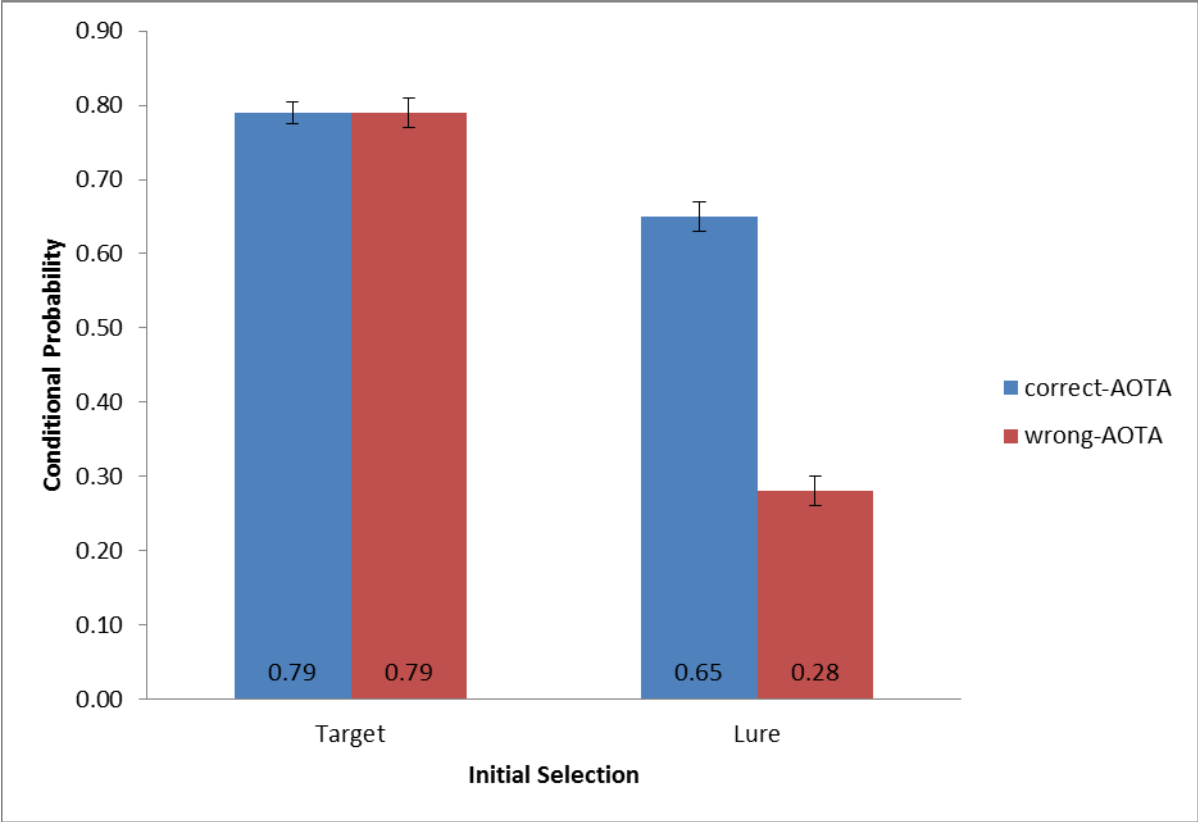


Figure 3. Conditional probabilities of accuracy on the final test for initial target retrieval and initial lure retrieval

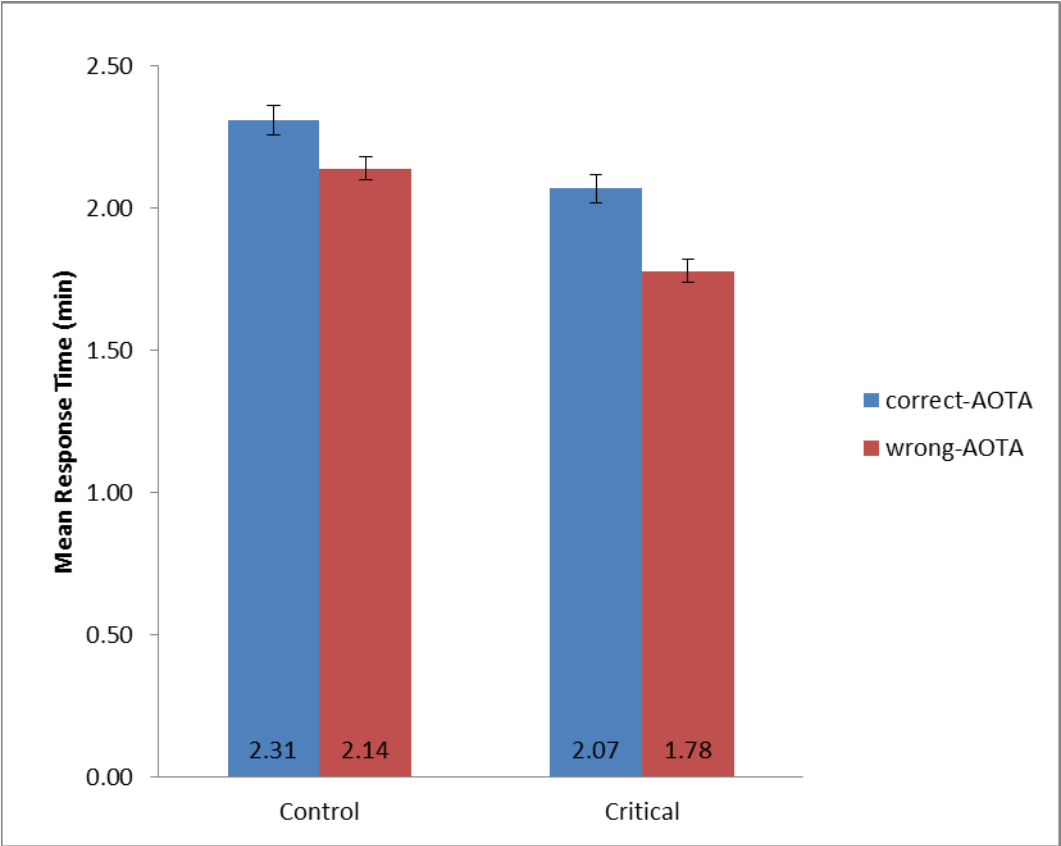


Figure 4. Mean response time (min) on the final test as a function of AOTA type and item type. Bars represent standard error.

Appendix A
GRE and TOEFL Reading Sample Used by Roediger and Marsh (2005)

Bats

Mammals of the order Chiroptera are only ones that have forelimbs modified as wings and hence are capable of true flight. Other mammals, such as the so-called flying lemurs and the flying squirrels, glide but do not fly. Bats, however, are so highly specialized as flying machines that locomotion by other means is accomplished with difficulty.

Bats are worldwide in their distribution in the temperate and tropical regions of the world and sometimes occur on remote oceanic islands that are uninhabited by other native mammals. On New Zealand, a pair of ancient continental islands, that is, islands once long ago connected with a mainland mass, two bats are the only mammals occurring naturally, everything else having been introduced by man. The order is divided into two suborders, the Megachiroptera and the Microchiroptera. The former contains only one family, the fruit-eating Pteropodidae of the Old World that is distinguished by the fact that most of its genera have a claw on the second digit, whereas in the Microchiroptera only the thumb bears a claw. The subordinal names are misleading because some of the Megachiroptera are smaller than some of the Microchiroptera.

Bats of the U.S. Gulf Coast are almost exclusively insect eaters, but elsewhere in the world there are bats whose principal diet is the pollen and nectar of flowers, fruit and vegetable matter, fish and other vertebrates such as small mammals, or blood obtained from a living victim.

The last group consists of the vampires of the American tropics that make a small incision in the skin, usually on the neck and shoulders or on the ears or toes, often while their prey is asleep, and then lap the blood that flows from the wound.

Appendix B
Sample Intervening MCQ Test

1. Bats are classified as part of which of the following mammalian order?
 - a) **chiroptera**
 - b) insectivora
 - c) marsupialia
 - d) Primata
 - e) AOTA

2. Where do the vampire bats make incision?
 - a) toes
 - b) shoulder
 - c) ears
 - d) neck
 - e) **AOTA**

3. What includes bats principle diet?
 - a) fruit
 - b) vegetable
 - c) fish
 - d) pollen
 - e) **AOTA**

4. Which of the following type of food is eaten by bats found on the U.S. Gulf Coast?
 - a) grasses
 - b) berries

- c) blood
- d) insects**
- e) AOTA

Sample Final Test MCQ Test

1. Bats are classified as part of which of the following mammalian order?
 - a. insectivora
 - b. chiroptera**
 - c. marsupialia
 - d. primate

2. Where do the vampire bats make incision?
 - a. heads
 - b. eyes
 - c. toes**
 - d. knees

3. What includes bats principle diet?
 - a. shrimps
 - b. honey
 - c. grass
 - d. vegetable**

4. Which of the following type of food is eaten by bats found on the U.S. Gulf Coast?
 - a. insects**
 - b. berries
 - c. blood
 - d. grasses