Transfer-Appropriate Processing in the Context of Complex Visual Structures and the Testing Effect

Melissa Ann Rowan Rice

Tufts University

Author Note

Melissa Ann Rowan Rice, Cognitive Aging and Memory Lab, Department of Psychology, Tufts University

Thank you to Dr. Ayanna Thomas, Greg Hughes and Tori Tavormina.

Thank you to committee members Dr. Ayanna Thomas and Dr. Marianna Eddy.

This work was made possible by the Tufts University Undergraduate Research Fund.

Abstract

Research has revealed that repeated testing results in better long-term retention as compared to repeated studying under educationally relevant conditions. The present study aims to elucidate further the conditions under which the testing effect is maintained by using complex visual stimuli. Participants were taught 12 easy and 12 hard chemical molecule Lewis-dot structures and their names in an initial study session before engaging in either retrieval practice (repeated testing) or study practice (repeated studying). During the final test, all participants were presented with two types of cues: picture and word. When a picture cue was presented, participants retrieved the name of the presented molecule; when a word cue was presented, participants retrieved the corresponding molecular structure. Through a final test manipulation, we also explored transfer-appropriate processing (TAP) as an explanation of the testing effect. The TAP account predicts a larger testing effect when the type of retrieval process utilized during practice matches the type of retrieval process required on the final test (i.e. same cue in both practice and final tests). In order to investigate TAP, within the retrieval practice group, final test cues were either congruent (picture/picture or word/word in practice/final) or incongruent (picture/word or word/picture in practice/final). Consistent with our predictions, participants correctly recalled significantly more easy items than hard items. We also found a TAP effect: participants in the retrieval practice condition correctly recalled significantly more congruent items than incongruent items. Finally, we found an advantage of retrieval practice over study practice when participants recalled hard items.

Keywords: testing effect, visual learning, transfer-appropriate processing,

Transfer-Appropriate Processing in the Context of Complex Visual Structures and the Testing Effect

In an effort to establish efficacious and efficient learning methods, significant research has been done to understand the benefits of repeated testing. Compared to repeated studying, repeated testing has been found to enhance long-term retention of educationally relevant material. Through repeated tests, students engage in retrieval practice—the "act of calling information to mind rather than rereading it or hearing it" (Roediger & Butler, 2011)—which results in increased learning and stronger long-term retention relative to studying. This phenomenon has been termed the *testing effect*. Multiple studies have investigated the conditions under which the testing effect is most powerful (see Karpicke & Grimaldi, 2012, for a comprehensive review). The majority of these studies have utilized educationally relevant verbal material, such as foreign language vocabulary lists (Karpicke & Roediger, 2008) and prose or scientific passages (Roediger & Karpicke, 2006a; Dobson & Linderholm, 2015). A small number of studies have investigated the testing effect using visual stimuli, such as pictures (Erdelyi & Becker, 1974; Wheeler & Roediger, 1992) and face-name pairs (Carpenter & DeLosh, 2005), but participants recalled only verbal information during testing. Furthermore, only a few studies have investigated the testing effect with complex spatial stimuli, such as maps (Carpenter & Pashler, 2007) and three-dimensional objects located in a virtual environment (Carpenter & Kelly, 2012). In addition to exploring the conditions under which the testing effect benefits long-term learning, researchers have also investigated *transfer-appropriate processing* (TAP) as a potential framework through which the testing effect can be understood (Morris, Bransford, & Franks, 1977; Veltre, Cho & Neely, 2015). The TAP account predicts a larger testing effect when the type of retrieval process utilized during practice matches the type of retrieval process required on the final test (i.e. same cue in both practice and final tests). The present study contributes to the existing literature by investigating whether the testing effect facilitates the learning of complex visual stimuli and by further examining transfer-appropriate processing in the context of the testing effect.

The design of the present study builds on past testing effect verbal learning studies and simultaneously addresses some limitations of previous research. Our design matched standard testing effect experiments, which typically include an initial study phase, a review phase and a final test. For example, in one such experiment (Roediger & Karpicke, 2006), participants learned prose passages and then reviewed the information via either repeated study (SSSS), a single test (SSST) or repeated tests (STTT). On a final test one week later, participants in the STTT and SSST groups performed significantly better than those in the SSSS group. These findings have been replicated in numerous studies with similar designs and other verbal stimuli (Dobson & Linderholm, 2015; for a review, see Roediger, Putnam & Smith, 2011). Even when participants engage in note-taking as part of the restudy session, participants who engage in retrieval practice perform significantly better on final memory tests (McDaniel, Howard & Einstein, 2009). In line with this established methodology, participants in the present study engaged in an initial study phase, then three sessions of either studying or retrieval practice, and finally an immediate memory test. The testing effect has been found to enhance performance on memory tests given at least one day after initial studying; the present study is the first experiment in a planned program of research in which we will incorporate a longer retention interval in subsequent experiments.

While previous research has found the testing effect to be robust in educationally relevant conditions with complex information (Karpicke & Aue, 2015), there is an obvious lack of exploration into whether it holds true with complex visual learning materials. In two studies that included visual stimuli as learning material (Erdelyi & Becker, 1974; Wheeler & Roediger, 1992), participants were presented with pictures of common items and words. However, during the practice and final memory tests, regardless of whether the participants recalled a word or a picture, they were instructed to write down the word (either the word item or the name of the item depicted in the picture). This is a problematic study design for exploring the effect of testing on picture recall because the participants recalled verbal information and were not forced to reproduce the images themselves. Our study addresses this problem by using complex visual structures that do not require artistic drawing ability to reproduce.

In a related study to the present experiment, Carpenter and Pashler (2007) used maps as stimuli in a testing effect study to explore whether testing can enhance learning of visuospatial information. In a within-subjects design, participants learned two maps, one by conventional repeated studying and the other by repeated testing. When learning the map through repeated testing, participants were presented with the map with one feature missing and had to recall the missing feature and its location. During studying, participants were simply presented with the complete map. During the final test, participants had to draw both maps completely. The final test occurred 30 minutes after initial studying. In line with previous testing effect studies on verbal learning, the repeated testing condition was significantly more beneficial for map retrieval than the study condition. These results suggest that repeated testing enhances recall of complex visuospatial information. Our study aims to extend these findings by also using complex visual stimuli.

Our aim in this experiment was to investigate whether the benefits of retrieval practice extend to the learning of complex visual structures and to explore TAP as a potential explanation for the testing effect. Participants were presented with 12 easy and 12 hard chemical molecules in the form of named Lewis-dot structures. As in previous testing effect studies, participants then engaged in either study practice or retrieval practice, immediately followed by final memory test. During the retrieval practice and final tests, participants were presented with two types of cues: picture and word. When a picture cue was presented, participants retrieved the name of the presented molecule; when a word cue was presented, participants retrieved the corresponding molecular structure. In order to investigate TAP, within the retrieval practice group final test cues were either congruent (picture/picture or word/word in practice/final) or incongruent (picture/word or word/picture in practice/final). As outlined above, our design is informed by previous testing effect studies, their limitations, and the possibility of using TAP as an explanation for the benefits of retrieval practice.

Based on previous research, we predicted:

1. Better recall of easy items over hard items
2. An advantage of retrieval practice over study practice
3. Better recall of congruent items over incongruent items within the retrieval practice group

We did not have a clear prediction about differences in recall accuracy between picture cues and word cues.

Method

Participants

Forty-eight undergraduate students at Tufts University in Medford, MA, participated for course credit or $20. Participants averaged 19 years of age. No participants had taken any college-level chemistry classes prior to or at the time the experiment. Participants were tested in groups of one to eight people at a time during one session. Each session was randomly determined to be repeated study or retrieval practice.

Materials

Twenty-four chemical molecules were chosen randomly and then converted into Lewis dot structures via the Wolfram-Alpha "Lewis structure" online widget (Lewis structure, 2015). A Lewis dot structure depicts a molecule in a two-dimensional graphical display. Molecules with less than nine atoms were labeled "easy"; molecules with more than nine but less than fourteen atoms were labeled "hard." Twelve easy and twelve hard molecules were chosen for this study. The images were converted to black-and-white, and all electron dots were erased so that each image showed only the atoms and bonds of each structure. Each molecule had three files: Pair—the Lewis dot structure with the molecule name added to the center-bottom of the structure; Picture—the unlabeled Lewis dot structure; Word—the molecule name. The files matched in size and font. Examples of chosen molecules are displayed in Figure 1.

The participants used 24-page booklets made from standard 8.5 x 11 inch white printer paper. Participants used a new blank booklet in each phase of the experiment, totaling five booklets: one for the study phase, three for the encoding phase, and one for the final test phase. The booklets were labeled according to experiment phase and subject number.

Procedure

Participants were tested in groups of one to eight people at a time. They were seated at a conference table with no physical dividers between them. The conference table was two to three feet away from a pull-down projector screen mounted on the wall. A standard classroom projector was connected to a laptop and used to display the experiment on the large screen.

Participants were instructed that they were going to engage in a memory test. They were told that they would study images of molecules in a series of study phases and would later have to recall missing information when cued (i.e. draw the structure when presented with the molecule name, or write the name when presented with the structure). The participants then completed a practice trial to get familiar with the study and test formats. Participants were given a four-page booklet identical to the experimental test booklets. The practice trial included a short study phase of two molecules (Pair files), and then a test phase. The instructions and the display time of the items were identical to the real experiment (described in detail below). The molecules used during the practice trial were not included in the experimental period.

At the start of the experimental period, all participants engaged in an identical initial study phase. During the initial study phase, all twenty-four molecules (Pair files) were displayed one at a time in the center of the screen in a random order. The participants were instructed to copy in their booklets each molecule structure and its name to ensure effective studying. Each molecule was displayed for thirty seconds and separated by a one-second long inter-stimulus interval (ISI). The ISI was indicated by a fully opaque, bright red screen. Every time the ISI occurred, participants flipped the page in their booklets to ensure that every molecule was on its own page. Participants in the study practice group continued with three more study phases identical to the initial phase. After each study phase, participants' booklets were collected and replaced by blank booklets. Participants were never given more than one booklet at a time.

Instead of repeated study, participants in the retrieval practice group continued with three study phases that resembled the final test phase. The retrieval practice phases were identical in duration to the study practice group's study phases. During retrieval practice, participants were shown a cue—either the Picture or Word file—for each molecule. When shown a Picture cue, participants retrieved the corresponding Word (the name of the molecule); when shown a Word cue, participants retrieved the corresponding Picture (the molecular structure). Each cue was displayed for twenty seconds and then the complete labeled structure (Pair) was displayed for ten seconds of feedback. A red screen (ISI) separated each display. The order of displayed molecules was blocked by cue type (Picture or Word). In each block, six easy molecules and six hard molecules were shown in random order. The block order was switched in each phase (e.g. Picture-Word, Word-Picture, Picture-Word), but the twelve molecules per block remained the same in each phase (e.g. easy 1-6 and hard 1-6 in Picture block, easy 7-12 and hard 7-12 in Word block). Four counterbalances ensured that every Picture and Word file was alternated equally.

All participants took an immediate final test. The final test was the same for both repeated study and retrieval practice groups. Twenty-four cues (Picture or Word) were presented one at a time for 30 seconds each with a one-second ISI in between. The twenty-four images were displayed in a random order. Half of the images were Picture cues (the participants had to write down the molecule name) and half were Word cues (the participants had to draw the molecular structure). The molecules were counterbalanced so that the difficulty of the molecules was distributed evenly (e.g. half of the Picture cues were easy, etc.). Within the retrieval practice group, every final test cue either matched the cue presented during retrieval practice trials (i.e. for a given molecule, the Picture cue was presented during both practice and final test or the Word cue was presented during both practice and final test) or did not match the cue presented during retrieval practice trials (i.e. for a given molecule, the Picture cue was presented during practice, but the Word cue was presented during the final test, or vice versa). Therefore, within the retrieval practice group, every final test cue was either congruent (matched) or incongruent (unmatched). Examples of cues presented in practice and final tests are shown in Figure 2. Congruency was distributed evenly (e.g., half of the easy Picture cues were congruent, etc.). Participants recorded their answers in their test booklets. Even if they could not recall the missing information, participants flipped the page at the ISI to ensure that their answers matched the display order. Prior to beginning the test, participants were encouraged to write down as much as they could remember on each trial.

Results

Scoring Procedure

The scoring of final tests was split between two scorers following the same scoring protocol. For both pictures and words, a binary scoring method was used: a recalled item was either correct or incorrect. A recalled picture item was correct if and only if 100% of all atoms and bonds were in the correct places. A recalled word item was correct if and only if the word was spelled 100% correctly. By applying a strict scoring procedure to all types of recalled items, we were able to compare binary scores between target types. The following analyses do not include scores from practice tests during retrieval practice because those tests have not yet been scored.

Analyses

A 2 (type of practice: study practice, retrieval practice) x 2 (item difficulty: easy, difficult) x 2 (cue type: picture, word) analysis of variance (ANOVA) was conducted on participants' final test scores. Three analyses were conducted: the first test compared congruent items in the retrieval practice group to all items in the study practice group; the second test compared incongruent items in the retrieval practice group to all items in the study practice group; the third test compared all items in the retrieval practice group to all items in the study practice group. Finally, a 2 (item difficulty: easy, difficult) x 2 (cue type: picture, word) x 2 (congruency: congruent, incongruent) analysis of variance (ANOVA) was conducted on the within-subjects manipulations in the retrieval practice group.

Analysis 1: Retrieval practice (congruent items) vs. Study Practice

Consistent with our predictions, participants performed significantly better on easy items (M = .42) than hard items (M = .17), $F(1, 50) = 76.942, p < .001, n^2p = .606.$ A significant interaction between difficulty and group, $F(1, 50) = 6.321, p = .015, n^2p = .112,$ revealed that retrieval practice (M = .22) lead to better final test performance on hard items than study practice (M = .13); when participants had to recall easy items, there was no difference between the two types of practice. Furthermore, we found a significant interaction between difficulty and cue type, $F(1, 50) = 7.313, p = .009, n^2p = .128$. When recalling hard items, participants performed better on picture cues (recalling the word) than on word cues (recalling the picture); when recalling easy items, pictures (M = .45) were recalled more accurately than words (M = .40). Finally, we found a marginally significant interaction between cue type and group, $F(1, 50) = 3.97, p = .052, n^2p = .074$, such that an advantage of retrieval practice was observed: retrieval practice led to better recall of words (M = .35) compared to study practice (M = .27).

Analysis 2: Retrieval practice (incongruent items) vs. Study Practice

This analysis revealed that participants recalled significantly more easy items (M = .39) than hard items (M = .12), $F(1, 50) = 98.187, p < .001, n^2p = .663.$ While there was no difference between groups during recall of hard items, an advantage of study practice was observed during recall of easy items (M = .45) over retrieval practice (M = .33). This indicates a transfer-appropriate processing effect; the benefit of retrieval practice is diminished when practice processes and final test processes are incongruent. This TAP effect is shown via a comparison of Figures 3 and 4, which display the interactions between difficulty and group from Analysis 1 and 2, respectively. Finally, we observed a significant interaction between difficulty and cue type: $F(1, 50) = 7.718$, $p = .008$, $n^2p = .134$. When recalling easy items, pictures were more accurately recalled (M = .43) than words (M = .35); when recalling hard items, words were more accurately recalled (M = .15) than pictures (M = .09).

Analysis 3: Retrieval practice (all items) vs. Study Practice

This analysis also revealed a significant effect of difficulty, $F(1, 50) = 146.255$, $p < .001$, $n^2p = .745$, such that easy items (M = .41) were recalled far more accurately than hard items (M = .15). As shown by the significant interaction between difficulty and group, $F(1, 50) = 8.917$, $p = .004$, $n^2p = .151$, we observed an advantage of retrieval practice when participants recalled hard items (M = .17) over study practice (M = .13). However, there was an advantage to study practice on easy items (M = .45) over retrieval practice (M = .36). This analysis also revealed an interaction between difficulty and cue type: $F(1, 50) = 11.836$, $p = .001$, $n^2p = .191$, such that participants recalled easy pictures (M = .44) better than easy words (M = .38) but recalled hard words (M = .18) better than hard pictures (M = .11), shown in Figure 5. Finally, we found a marginally significant interaction between cue type and group: $F(1, 50) = 3.463$, $p = .069$, $n^2p = .065$. This indicates a slight advantage of retrieval practice when participants recalled words (M = .29) over study practice (M = .27), but a slight advantage of study practice on picture recall (M = .30) over retrieval practice (M = .24).

Analysis 4: Within retrieval practice: difficulty x target type x congruency

Finally, we conducted a within-subjects ANOVA for the retrieval practice group. There were significant main effects of congruency, $F(1, 25) = 5.52$, $p = .027$, $n^2p = .18$, and difficulty, $F(1, 25) = 40.71$,

$p < .001$, $n^2p = .62$. In line with the TAP account, participants recalled more congruent items (M = .31) than incongruent items (M = .22). Participants also recalled more easy items (M = .34) than hard items (M = .20). There was no effect of cue type.

General Discussion

The present study provides a solid foundation for further exploration into the effects of repeated testing on complex visual material as compared to repeated studying. We predicted that participants would do better on easy items (Hypothesis 1), that there would be an effect of retrieval practice (Hypothesis 2), and an effect of congruent items within retrieval practice (Hypothesis 3). Our results overall were consistent with Hypothesis 1 and Hypothesis 3, and significant interactions revealed an advantage of retrieval practice under specific conditions, complicating but supporting Hypothesis 2. In the following discussion, we present possible explanations for our results.

Analysis 1—congruent retrieval practice items compared to study practice—shows an advantage of retrieval practice over study practice during recall of difficult items, which may be explained by the *desirable difficulty* framework. In a study investigating the desirable difficulty theory as it applies to the testing effect, Pyc and Rawson (2009) found that that increased performance on the final test is correlated with increased difficulty of retrieval practice. Further analysis of data from the retrieval practice trials is necessary to determine whether "successful but difficult processing [is] better for memory than successful but easier processing" (Pyc & Rawson, 2009), a result predicted by the desirable difficulty framework. Even without these data, our results are important because they contribute to the understanding of the factors that underlie the testing effect, and they corroborate the finding that retrieval practice can be used to improve memory of complex material (Karpicke & Aue, 2015).

Despite the advantage of retrieval practice during recall of hard items, we found no significant difference between study techniques on recall of easy items. This may be due to the fact that participants took a test only immediately following study sessions. Past research has found that the testing effect is particularly effective after a retention interval of 24 hours or more (see Veltre et al., 2015). The present study is the first experiment in a planned program to investigate the testing effect and complex visual materials; our follow-up study will include a 48-hour retention interval, and we predict the testing effect will be more evident on that delayed final memory test. Furthermore, we predict that by including a retention interval we will get a better idea of how the two study techniques differ when participants retrieve pictures as opposed to words. The marginally significant interaction between difficulty and group found in Analysis 1 shows an advantage of retrieval practice over study practice when participants recalled words, which has been found in numerous testing effect studies using verbal materials (Roediger et al., 2011) but no difference on picture retrieval between groups. Carpenter and Pashler's (2007) study showed that retrieval practice does benefit retention of visuospatial stimuli, suggesting that further research using Lewis-dot structures would show the same.

A comparison between Analysis 1 and Analysis 2 (see Figures 3 and 4) and results from Analysis 4 reveal a TAP effect as predicted. TAP predicts that matching cues on retrieval practice and final tests will result in stronger memory performance; our results show that participants recalled significantly more congruent items than incongruent items. (This analysis does not apply to study practice participants as they did not practice retrieval during review.) Furthermore, the benefit of retrieval practice on hard items was lost (shown by Analysis 2: incongruent retrieval practice items compared to study practice) when those items were incongruent, indicating that participants engaging in retrieval practice were put at a disadvantage when final test cues were incongruent to practice test cues. These results are important theoretically in that they provide evidence for the TAP framework.

Finally, the significant interaction between difficulty and cue type found in Analysis 3 has potential implications that must be discussed. Participants recalled easy pictures better than easy words, but recalled hard words better than hard pictures (shown in Figure 5). While this may indicate a difference in verbal and visual processing, a closer look at the stimuli used reveals possible verbal interference on easy items (i.e. "ethylene" and "ethane"). We must do further analysis on these particular items to make sure that there was not interference between the two items before claiming a difference between visual and verbal processing.

Our findings contribute to the existing literature on the testing effect because we found evidence for a TAP-potentiated advantage of retrieval practice on difficult items using novel complex visual stimuli. Our proposed future research will incorporate a retention interval into the design and eventually expand into using three-dimensional molecular structures as stimuli.

References

Carpenter, S. K., & Delosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology*, 19(5), 619-636. doi:10.1002/acp.1101

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14(3), 474-478.

Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review*, 19(3), 443-448. doi:10.3758/s13423-012-0221-2

Dobson, J. L., & Linderholm, T. (2014). Self-testing promotes superior retention of anatomy and physiology information. *Advances in Health Sciences Education*, 20(1), 149-161. doi:10.1007/s10459-014-9514-8

Erdelyi, M. H., & Becker, J. (1974). Hypermnesia for pictures. *Cognitive Psychology*, 6(1), 159-171. doi:10.1016/0010-0285(74)90008-5

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-Based Learning: A Perspective for Enhancing Meaningful Learning. Educational Psychology Review, 24(3), 401-418. doi:10.1007/s10648-012-9202-2

Karpicke, J. D., & Aue, W. R. (2015). The Testing Effect Is Alive and Well with Complex Materials. *Educational Psychology Review*, 27(2), 317-326. doi:10.1007/s10648-015-9309-3

Lewis structure. (2013). Retrieved April 15, 2017, from http://www.wolframalpha.com/widgets/view.jsp?id=689aa5a01c216d8b16ed0250cebdc702

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The Read-Recite-Review Study Strategy. *Psychological Science*, 20(4), 516-522. doi:10.1111/j.1467-9280.2009.02325.x

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519-533. doi:10.1016/s0022-5371(77)80016-9

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437-447. doi:10.1016/j.jml.2009.01.004

Roediger, H. L., I.,II, & Karpicke, J. D. (2006) Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science,* 17(3), 249-55.

Roediger, H. L., I.,II, & Butler, A. C. (2011) The Critical Role of Retrieval Practice in Long-Term Retention. *Trends in cognitive sciences,* 15(1), 20-7.

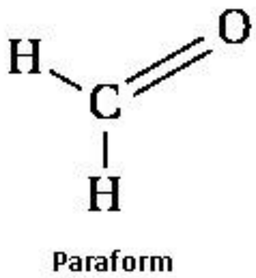
Roediger, H. L., Iii, Putnam, A. L., & Smith, M. A. (2011). Ten Benefits of Testing and Their Applications to Educational Practice. *Psychology of Learning and Motivation*, 1-36. doi:10.1016/b978-0-12-387691-1.00001-6

Veltre, M. T., Cho, K. W., & Neely, J. H. (2014). Transfer-appropriate processing in the testing effect. *Memory*, 23(8), 1229-1237. doi:10.1080/09658211.2014.970196

Weldon, M. S., Roediger, H. L., & Challis, B. H. (1989). The properties of retrieval cues constrain the picture superiority effect. *Memory & Cognition*, 17(1), 95-105. doi:10.3758/bf03199561

Wheeler, M. A., & Roediger, H. L. (1992). Disparate Effects of Repeated Testing: Reconciling Ballard's (1913) and Bartlett's (1932) Results. *Psychological Science*, 3(4), 240-245. doi:10.1111/j.1467-9280.1992.tb00036.x

Appendix A

A) 

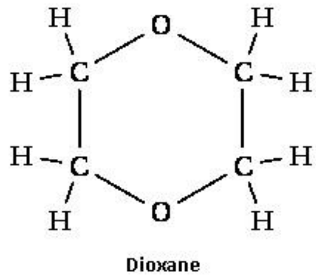
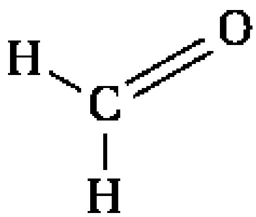
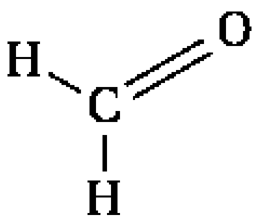
B) 

Figure 1. Examples of easy (A) and difficult (B) stimuli.

C) 

D)¶ 

E) 

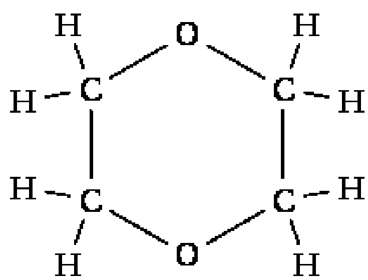
F) 

Figure 2. Examples of test cues. Item C is an example of a picture cue—participants retrieve the molecule name, in this case "paraform." Item D is an example of a word cue—participants retrieve the molecule structure (see item F). Given that C and D are retrieval practice test cues, items E and F are examples of congruent and incongruent final test items, respectively. Item E is congruent because participants were presented with the same cue (C) during practice; item F is incongruent because participants were presented with the corresponding cue (D) during practice.

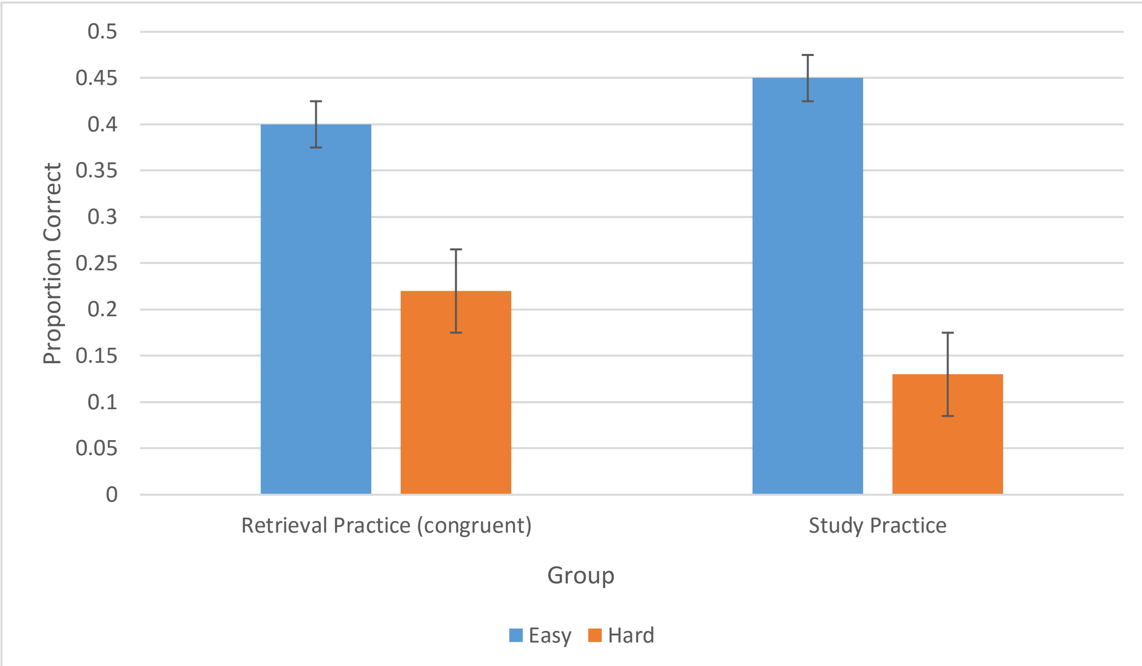


Figure 3. Average scores on easy and hard items between congruent retrieval practice items and study practice items. This figure shows the significant interaction between difficulty and group type. There is an advantage to retrieval practice as compared to study practice when participants recall hard items.

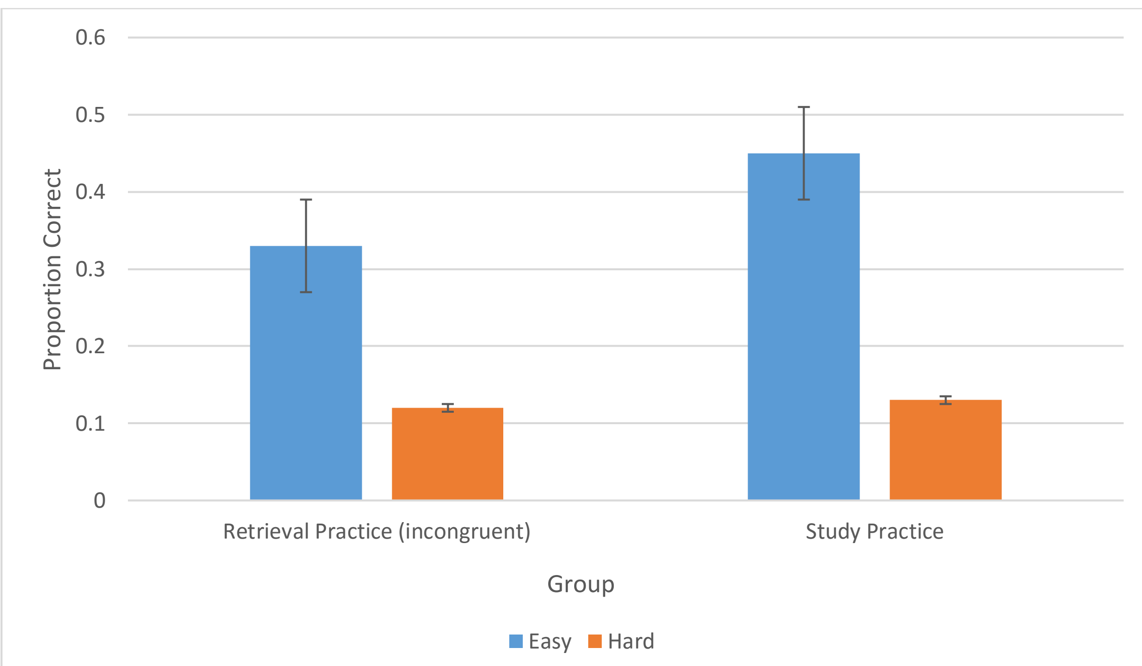


Figure 4. Average scores on easy and hard items between incongruent retrieval practice items and study practice items. This figure shows the significant interaction between difficulty and group type. There is an advantage to study practice on easy items when compared to incongruent retrieval practice easy items.

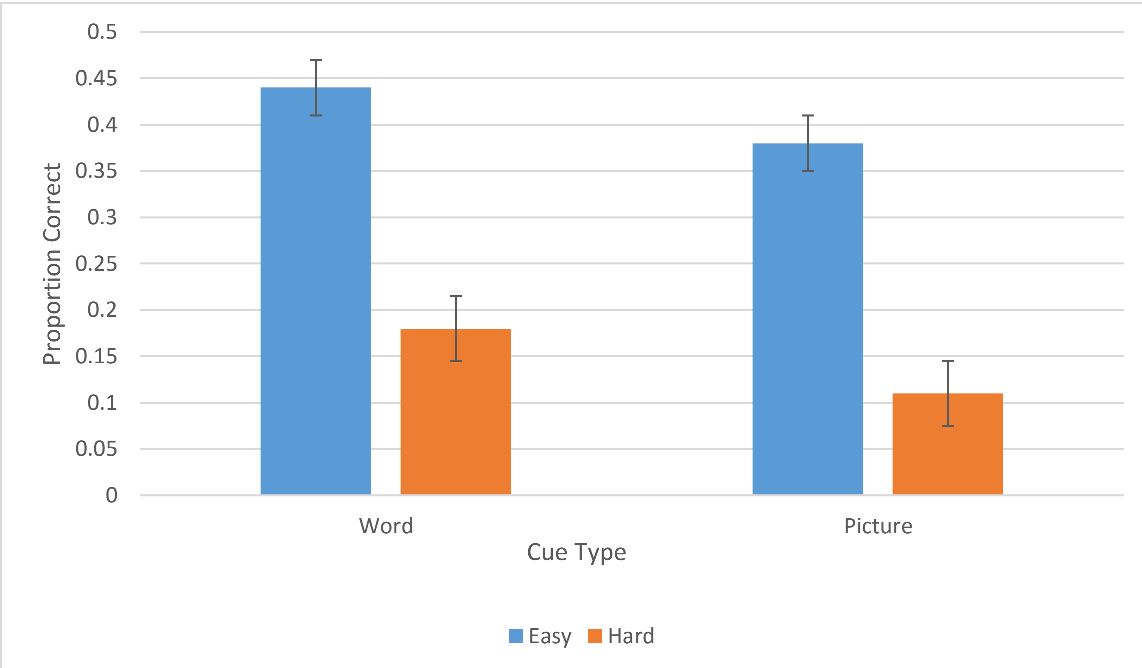


Figure 5. Average scores on easy and hard items across cue type when comparing all retrieval practice items and study practice items. This figure shows the significant interaction between cue type and difficulty.