What Types of Learning are Enhanced by a Cued Recall Test?

SHANA K. CARPENTER, HAROLD PASHLER, AND EDWARD VUL
University of California, San Diego

In 2 experiments we investigated what types of learning benefit from a cued recall test. After initial exposure to a word pair (A + B), subjects experienced an intervening cued recall test (A → ?) with feedback, or a re-study presentation (A → B). The final test could be cued recall in the same direction (A → ?) or the opposite direction (? → B), or free recall of just the cues (Recall As) or just the targets (Recall Bs). All final tests revealed a benefit of testing over re-studying. Tests produced a direct benefit for the information that was retrieved on the intervening test (B), but also "spilled over" to facilitate recall of information that was present on the test but was not retrieved (A). Both theoretical and practical implications are discussed.

What Types of Learning are Enhanced by a Cued Recall Test?

A memory test is commonly used to measure the accuracy or speed of memory. However, a memory test can also *modify* memory, sometimes in a beneficial way. For example, Duchastel (1981) showed that students remembered textbook information better if they completed test questions on the information, rather than engaged in an unrelated activity. Furthermore, a number of studies have found that testing is even more beneficial than additional study presentations (Carpenter & DeLosh, in press; Carrier & Pashler, 1992; Kuo & Hirshman, 1996, 1997; Wheeler, Ewers, & Buonanno, 2003). This benefit for tested over re-studied information is often referred to as the *testing effect* (see Dempster, 1996, for a review).

We can shed light on *why* the testing effect occurs by asking what types of learning benefit from testing. Are testing benefits confined to the very items that were retrieved on the test? Or do they also occur for items that were on the test but not retrieved? If the benefits are confined to the retrieved items, do they only show up when the final test is the same type as the intervening test? We examined these questions using cued recall (A → B). Previous research indicates that a cued recall test (A → ?) is more beneficial than re-study (A + B) when the final test is cued recall in the same direction (A → ?) (Carpenter & DeLosh, 2005; Carrier & Pashler, 1992; Cull, 2000; Izawa, 1969, 1992). However, do these benefits also occur when the final test is cued recall in the opposite direction (? → B), or free recall of just the targets (Recall Bs) or cues (Recall As)?

This question has clear practical implications. Many have argued that the testing effect may have important and unexploited educational potential (e.g., Chan, McDermott, & Roediger, 2005; Dempster, 1989, 1996; Glover, 1989; McDaniel & Fisher, 1991; Roediger & Karpicke, in press). Before concluding this, however, we must know whether these benefits occur for one sort of memory but not for another. For example, one's enthusiasm for testing as a way to enhance learning of the German-English correspondence Hund ↔ Dog would be tempered if a test (Hund → ?) enhanced forward recall but not backward recall (? → Dog). In the current study, we explored the breadth of the testing effect to find out when tests are beneficial, or conceivably harmful, relative to re-study opportunities.

Experiment 1

In Session 1, subjects were presented with 40 weakly related cue-target pairs. After a study presentation, subjects were then given an additional opportunity to learn each pair. This either took the form of an opportunity to re-study the pair (A + B) or a cued recall test (A → ?) which was immediately followed by presentation of the pair (A + B). The 2 types of additional learning opportunities are referred to here as *Study Trials* and *Test/Study Trials*, respectively. The duration of the 2 types of trials was always equated in the experiments presented here. The following day (Session 2), subjects completed 1 of 4 different types of final tests: cued recall in the same direction (A → ?) or

Corresponding Author: Shana K. Carpenter, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109. Phone: (858) 822-1840. Fax: (858) 534-7190 E-mail: scarpenter@ucsd.edu

opposite direction (? → B) as the Test/Study Trial of Day 1, or free recall over just the cues (Recall As), or just the targets (Recall Bs).

Method

Subjects. We recruited subjects from an online pool of individuals who volunteered to complete the experiment in exchange for enrollment in a drawing for cash prizes. Out of 365 subjects who began the experiment, 90 dropped out during Session 1, 49 during Session 2, and 50 failed to follow instructions (e.g., waited longer than 48 hours to complete Session 2). The remaining 176 subjects were randomly distributed across the 4 final test conditions: A → ? (n = 43), ? → B (n = 53), Recall As (n = 45), and Recall Bs (n = 35).

Internet testing allowed us to collect data from a larger and more demographically diverse group of subjects than would be possible with standard laboratory testing. Although web-based data collection has only recently become common, parallel patterns of results have been obtained in numerous laboratory- and web-based experiments both in our own research and in others' (e.g., Birnbaum, 1999; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Reips, 2002). Analysis of subjects' reports about what environment they were in while participating provided further reassurance that even those participating in semi-public environments did not show any detectable decrements or changes in performance (see results). While web-based studies have higher dropout rates than laboratory studies, the critical manipulation of Study versus Test/Study was carried out within subjects to avoid any possibility of differential drop-out effects.

Materials. From Wilson's (1988) database, we obtained 80 nouns that were 5-7 letters, 1-3 syllables, and high in concreteness (400-700) and word frequency (at least 30 per million). Free-association norms (Nelson, McEvoy, and Schreiber 1998) were used to create 40 weakly associated pairs of similar forward and backward strength (see Appendix). Each word in a pair was randomly assigned to be cue or target for each subject.

Design and Procedure. Subjects first read instructions and answered several demographic questions, including what type of environment they were in while doing the experiment (e.g., at home, in an office, in an Internet café, in a library). The experiment began with the presentation of the 40 word pairs, one at a time, for 6 seconds each. The cue appeared on the left and the target on the right, each in separate boxes with the labels "cue" and "target" above them.

We used a 2 x 2 x 2 mixed design. The within-subjects factor (test condition: test/study vs. study) was manipulated across items during Session 1. First all 40 word pairs were presented, then subjects completed a Test/Study Trial on 20 of the word pairs. During a Test/Study Trial, subjects were instructed to try to covertly retrieve the target within 4 seconds while the "cue" box displayed the cue and the "target" box was blank. After 4 seconds, the target appeared and both items remained present in their respective boxes for an additional 2 seconds. For the other 20 word pairs, subjects completed a Study Trial, in which they were given an additional study opportunity to view the cue and target in their respective boxes for 6 seconds. Session 1 was complete after all 40 word pairs were presented in either a Test/Study Trial or a Study Trial. The assignment of items to test condition, and their order of presentation, was random for each subject.

The 2 between-subjects factors (item retrieved on final test: cue vs. target, and type of final test: cued recall vs. free recall) were manipulated during Session 2, which subjects could access via e-mail between 18 and 48 hours following Session 1. A combination of the 2 factors yielded 4 final test conditions, and subjects were randomly assigned to one of these. They were instructed to either: (1) type the correct target when given the cue (A → ?), (2) type the correct cue when given the target (? → B), (3) type all of the targets they could remember (Recall Bs), or (4) type all of the cues they could remember (Recall As). No time limit was imposed, and no feedback was provided. Session 2 was completed when subjects typed an answer to all 40 items for the cued recall tests, or when they clicked a button marked "finish" to indicate they could no longer remember any items for the free recall tests.

Results and Discussion

Most subjects (72% in Session 1, and 76% in Session 2) reported that they performed the experiment while in a room alone, and the rest were fairly evenly distributed among the other environments. Environment did not significantly affect final test accuracy, nor did it interact with any variables.

Test/Study Trials produced higher final test accuracy (40% overall) than Study Trials (30% overall). The testing benefit appeared regardless of the nature of the final test (see Figure 1). When the final test required cued recall in the same direction (A → ?) or the opposite direction (? → B), the testing benefit was 14%. When the final test required free recall of the targets (Recall Bs) or the cues (Recall As), the testing benefit was 8% and 6%, respectively. The significance of these effects was examined in a 2 (test condition) x 2 (item retrieved on final test) x 2 (type of final test) Analysis of Variance (ANOVA). There was a main effect of test condition [$F(1, 172) = 101.58$, $p < .001$, $MSE = .009$] but no main effect for item retrieved on the final test ($F = 3.10$), and no interaction ($F = .15$).

There were 3 additional significant effects. The first was a main effect for type of final test indicating that, unsurprisingly, cued recall was easier than free recall [$F(1, 172) = 244.76$, $p < .001$, $MSE = .083$]. The second was an interaction between test condition and type of final test, indicating that the testing effect was larger for cued recall than free recall [$F(1, 172) = 11.99$, $p < .005$, $MSE = .009$], probably because free recall was subject to floor effects. The third was an interaction between type of final test and item retrieved on the final test, indicating that cued recall showed an advantage of targets (A → ?) over cues (? → B), whereas free recall did not (Recall As = Recall Bs) [$F(1, 172) = 3.975$, $p < .05$, $MSE = .083$].[1]

In sum, the key results from Experiment 1 are that the benefits of the intervening cued recall test occurred whether the final test required cued recall in the same or opposite direction as the intervening test, and even when the final test required free recall of the items for which retrieval was required on the intervening test (targets), and free recall of the items for which retrieval was not required (cues).

Experiment 2

The methodology of Experiment 1 had one possible disadvantage. During a Study Trial, subjects were given 6 seconds to just read the word pair again. During that time, they may have thought about other, previously presented word pairs. Subjects sometimes use the time available during presentation of one item to think about a previous difficult-to-learn item (Slamecka & Katsaiti, 1987), and the Test/Study items could be considered more difficult, because it is harder to access information through retrieval than mere presentation.

When Test/Study and Study pairs are presented in random order, as in Experiment 1 (…Test/Study → Study → Study → Test/Study → Study → Test/Study…), a previous Test/Study can be easily retrieved during a Study. This is harder to do when the order is blocked so that all Test/Study come before Study (…Test/Study → Test/Study → Test/Study → Study → Study → pure ST…), and it is impossible to do when the order is blocked so that all Study come before all Test/Study (…Study → Study → Study → Test/Study → Test/Study → Test/Study…).

Thus, in Experiment 2, subjects received 1 block of 20 Study followed by 1 block of 20 Test/Study, or vice versa. The order of the blocks, and the items within blocks, was random for each subject. In all other respects, Experiment 2 was identical to Experiment 1. We recruited new subjects from the same pool as before. Out of 177 who began the experiment, 43 dropped out during Session 1, 24 during Session 2, and 28 failed to follow instructions. The remaining 82 subjects were randomly distributed across the final test conditions: A → ? (n = 19), ? → B (n = 19), Recall As (n = 18), and Recall Bs (n = 26).

Results and Discussion

Most subjects (74% in both sessions) completed the experiment while in a room alone, and the rest were fairly evenly distributed among the other environments. As in Experiment 1, environment did not significantly affect final test accuracy, nor did it interact with any variables.

Test/Study produced higher final test accuracy (40% overall) than Study (31% overall). The testing benefit appeared regardless of the nature of the final test (see Figure 1). When the final test required cued recall in the same direction (A → ?) or the opposite direction (? → B), the testing benefit was 14% and 9%, respectively. When the final test required free recall of the targets (Recall Bs) or the cues (Recall As), the benefit was 4% and 8%, respectively.

Experiment 2 replicated the same basic pattern of ANOVA results from Experiment 1: a significant main effect for test condition [$F(1, 78) = 30.80$, $p < .001$, $MSE = .011$], but no main effect for item retrieved on the final test ($F = .48$), and no interaction ($F = .04$). Two other significant effects were found: a main effect for type of final test, reflecting the fact that cued recall was easier than free recall [$F(1, 78) = 111.25$, $p < .001$, $MSE = .076$], and an interaction between type of final test and test condition, indicating that the testing effect was larger for cued recall than free recall [$F(1, 78) = 4.01$, $p < .05$, $MSE = .011$], probably because free recall was subject to floor effects.

The blocked order of Test/Study vs. Study made it unlikely that cues were retrieved during the intervening test. Nonetheless, Experiment 2 still showed a testing effect for final tests that were either the same or opposite direction as the intervening test, and for items that were required to be retrieved (targets) and items that were not required to be retrieved (cues).

General Discussion

In 2 similar experiments, what we refer to as a Test/Study Trial—an intervening cued recall test (A → ?) followed by re-presentation of the word pair (A + B)—enhanced retention more than did a comparable amount of time provided for pure study (A + B). This held true whether retention was tested for cued recall in the same (A → ?) or opposite (? → B) direction as compared to the intervening test, and free recall of either the targets (Recall Bs) or the cues (Recall As). The significant testing effect in the same direction replicates previous reports (Carpenter & DeLosh, 2005; Carrier & Pashler, 1992; Cull, 2000; Izawa, 1969, 1992). However, the current study extends these findings by showing that the testing effect is not specific to the items for which retrieval was required on the intervening test, nor to the type of testing employed.

Theoretical Implications

Additional Study Time. We obtained no evidence that tested items benefit simply because they receive more study time than non-tested items. First, an intervening Test/Study Trial was more beneficial than a Study Trial, even though the cue and target were presented together for more time in a Study Trial (6 seconds) than in a Test/Study Trial (2 seconds). Second, we obtained a significant testing effect whether Test/Study and Study Trials were inter-mixed (Experiment 1) or blocked (Experiment 2). It therefore seems unlikely that a Test/Study Trial produces superior learning because it "steals" study time away from a Study Trial. Our results are consistent with past studies that obtained the testing effect using blocked lists (e.g., Carrier & Pashler, 1992) and between-subjects manipulations of test vs. re-study (Wenger, Thompson, & Bartling, 1980).

Transfer-Appropriate Processing. The processes required by an intervening test and final test are more similar, relative to the processes required by an intervening study opportunity and final test. According to a transfer-appropriate processing (TAP) view (e.g., Morris, Bransford, & Franks, 1977), tests could benefit learning simply because they provide practice at the relevant aspects of the task that are needed for the final test. Some studies have supported this notion by showing that intervening tests are more effective if they are more similar to the final test (e.g., McDaniel & Fisher, 1991; McDaniel, Kowitz, & Dunay, 1989). However, in the current experiments, the intervening test always required recall in one direction (A → ?), and we observed a testing effect whether the final test required recall in the same (A → ?) or opposite (? → B) direction, or recall of just the targets (Recall Bs) or the cues (Recall As). Consistent with past studies and contrary to the TAP view, we found that an intervening test was beneficial to retention even if the final test was of a different type (Carpenter & DeLosh, in press; Glover, 1989; Kang, McDermott, & Roediger, 2005).

Error Correction Learning. Carrier and Pashler (1992) proposed an explanation for the testing effect based on error-correction learning models (e.g., McClelland & Rumelhart, 1986). According to this view, the association between 2 items (A → B) is learned by adjusting connections in a network in a way that minimizes the error in producing B from A. If retrieval of B is required (A → ?) as in a Test/Study Trial, then learning occurs by comparing one's actual response (B') with the desired response (B) to determine how much adjustment is necessary (see also Mozer, Howe, & Pashler, 2004). When both items are present (A + B) as in a pure Study trial, learning is impoverished because having B available makes it harder for the system to ascertain what response it would produce on its own, thus interfering with the calculation of appropriate weight changes. It is not clear how this hypothesis would account for the occurrence of an advantage for Test/Study Trials when the final test runs in the opposite direction (? → B), since the to-be-retrieved item A was never produced on the intervening test.

Practical Implications

The generality of the testing effect suggests that tests have great potential to enhance learning in practical domains. Specifically, the use of flashcards likely improves recallability not only in the direction that was practiced (e.g., German-English vocabulary Hund → Dog), but also in the direction that was not practiced (Dog → Hund). Tests might also be useful in improving patients' recall of medical information, which is frequently misremembered (e.g., Kessels, 2003). For example, a patient's memory for symptoms *and* medications may be improved by attempting to recall what medication to take when experiencing specific symptoms.

By using tests with feedback regularly in lieu of re-studying the same material over again, it appears there is much to be gained and little, if anything, to be lost. A promising direction for further research is to explore how the testing effect might be obtained for knowledge that is more complex and structured than the paired associate information examined in this and most other studies of testing effects.

Author Note

Shana K. Carpenter, Harold Pashler, and Edward Vul, Department of Psychology, University of California at San Diego. Edward Vul is now at the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology. This work was supported by the Institute of Education Sciences (US Department of Education, Grant R305H040108). We thank Matt Bielich for his programming expertise in both experiments.

Footnotes

[1] The advantage for targets over cues does not seem inconsistent with the principle of associative symmetry (for an excellent review, see Kahana, 2002), as this advantage was modest in Experiment 1 (p = .048) and non-existent in Experiment 2. Rather, the advantage seems to be influenced by random error combined with the fact that subjects were not aware of what type of final test they would receive, and thus could have reasonably expected another test in the same direction instead of the opposite direction.

References

Birnbaum, M. (1999). Testing critical properties of decision making on the internet. *Psychological Science*, 10, 399-407.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology*, 19, 619-636.

Carpenter, S. K., & DeLosh, E. L. (in press). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633-642.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. III. (2005). *Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material*. Manuscript submitted for publication.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215-235.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review*, 4, 309-330.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork and R. A. Bjork (Eds.), *Handbook of perception & cognition: Memory* (pp. 317-344). San Diego, CA: Academic Press.

Duchastel, P. C. (1981). Retention of prose following testing with different types of tests. *Contemporary Educational Psychology*, 6, 217-226.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392-399.

Izawa, C. (1969). Comparison of reinforcement and test trials in paired-associate learning. *Journal of Experimental Psychology*, 81, 600-603.

Izawa, C. (1992). Test trials contributions to optimization of learning processes: Study/test trials interactions. In A. F. Healy & S. M. Kosslyn (Eds.), *Essays in honor of William K. Estes, Vol. 1: From learning theory to connectionist theory* (pp. 1-33). Hillsdale, NJ: Erlbaum.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, 30, 823-840.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. III. (2005, May). *Testing enhances memory retention, but which test format is better?* Poster session presented at the annual meeting of the American Psychological Society, Los Angeles, CA.

Kessels, R. P. C. (2003). Patients' memory for medical information. *Journal of the Royal Society of Medicine*, 96, 219-222.

Krantz, J. H., & Dalal, R. (2000). Validity of web-based psychological research. In M. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. 35-60). San Diego, CA: Academic Press.

Kuo, T. M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, 109, 451-464.

Kuo, T. M., & Hirshman, E. (1997). The role of distinctive perceptual information in memory: Studies of the testing effect. *Journal of Memory & Language*, 36, 188-201.

McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 2: Psychological & biological models* (pp. 170-215). Cambridge, MA: MIT Press.

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology*, 16, 192-201.

McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, 17, 423-434.

McGraw, T. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science*, 11, 502-506.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, 16, 519-533.

Mozer, M. C., Howe, M., & Pashler, H. (2004). Using testing to enhance learning: A comparison of two hypotheses. *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 975-980). Hillsdale, NJ: Erlbaum Associates.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://www.usf.edu/FreeAssociation/

Reips, U-D. (2002). Standards for internet experimenting. *Experimental Psychology*, 49, 243-256.

Roediger, H. L. III., & Karpicke, J. D. (in press). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, 26, 589-607.

Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 135-144.

Wheeler, M. A., Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11, 571-580.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioural Research Methods, Instruments, & Computers*, 20, 6-11.

Appendix

Item Pairs		Forward Strength	Backward Strength	Absolute Value Difference
angle	corner	.020	.029	.009
author	poet	.028	.035	.007
beach	blanket	.012	.016	.004
block	street	.040	.019	.021
chain	fence	.022	.031	.009
child	mother	.030	.010	.020
cloth	table	.012	.026	.014
coffee	morning	.025	.034	.009
college	student	.035	.046	.011
curve	shape	.018	.011	.007
engine	machine	.033	.027	.006
factory	product	.020	.028	.008
frame	window	.014	.013	.001
group	meeting	.027	.041	.014
guard	prison	.024	.020	.004

lunch	supper	.019	.028	.009
master	owner	.010	.028	.018
nation	state	.042	.055	.013
native	foreign	.056	.031	.025
nature	trail	.023	.012	.011
novel	story	.034	.034	0
object	symbol	.014	.021	.007
office	doctor	.014	.010	.004
paint	picture	.036	.031	.005
pencil	point	.021	.073	.052
people	world	.014	.030	.016
quarter	dollar	.061	.027	.034
range	rifle	.015	.028	.013
report	weather	.015	.024	.009
sheet	cover	.021	.053	.032
slave	worker	.069	.062	.007
smile	teeth	.061	.042	.019
sound	speaker	.024	.027	.003
station	radio	.067	.095	.028
stick	branch	.067	.047	.020
store	general	.016	.028	.012
taste	touch	.016	.012	.004
throat	voice	.039	.020	.019
train	plane	.051	.049	.002
vehicle	truck	.013	.014	.001
	Mean	.029	.032	.002

Figure Caption

Figure 1. Percent of items recalled on the final tests. Items experienced an intervening cued recall test with feedback (Test/Study) vs. a re-study opportunity (Study). Test/Study items were recalled better than Study items whether the final test required cued recall in the same direction (A - ?) or opposite direction (? - B) as the intervening test, or free recall of just the cues (Recall As) or just the targets (Recall Bs). Error bars represent standard errors.

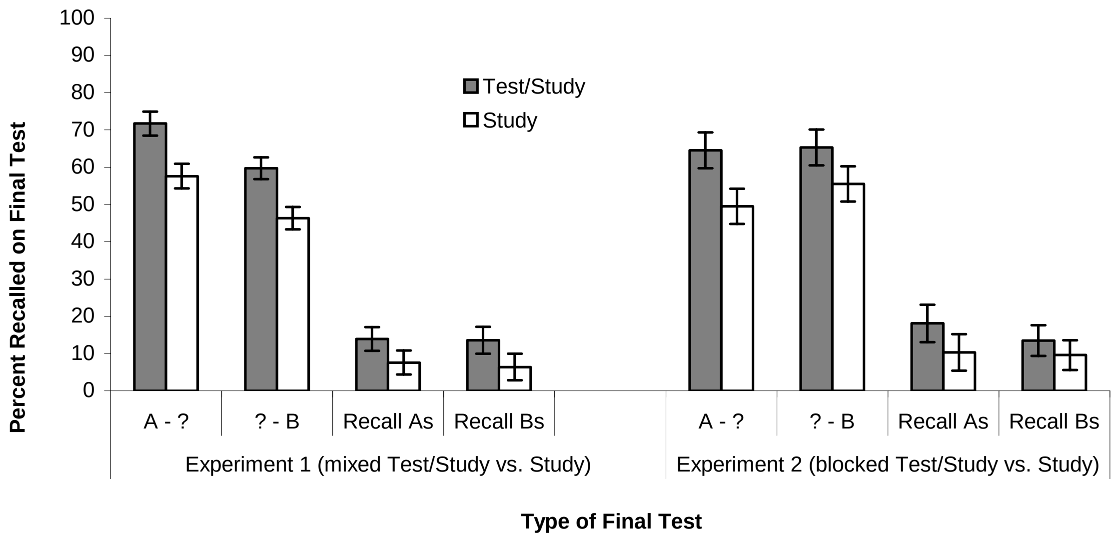

Figure 1