Investigating the Semantic Activation Theory of the Pretesting Effect

Meriç Karaoğuz, Jeffrey D. Karpicke, & Michelle E. Coverdale

Purdue University

Abstract

In word-pair experiments, when people are given a cue and attempt to guess the target (e.g., *tide - ?*) prior to studying the cue and target (e.g., *tide – beach*), their memory is better on a final test than if they only studied the pair. This phenomenon is known as the "pretesting effect". One theory proposes that during a pretest, people search for and activate semantically related words, and partial semantic activation enhances encoding. We tested this prediction by measuring priming on a lexical decision task, a well-established way to assess semantic activation. After taking a pretest or studying each item, subjects completed a lexical decision trial. Although we replicated the pretesting effect on a final memory test, we did not observe semantic priming on the lexical decision task following a pretest. In fact, priming was greater in the study-only condition. These results challenge semantic activation theories of the pretesting effect.

Introduction

Pretesting refers to the finding that people perform better on a final test when they are given a test (a “pretest”) prior to initial learning (Kornell, et al., 2009). In other words, testing people on a cue word they have not yet learned, prior to learning a cue-target pair, improves memory on a final cued-recall test compared to only studying the cue-target pair. For example, showing a participant the cue word “suds” and asking them to guess a related word that it might be paired with before showing them the actual cue and target pair, “suds-bath” in this case, leads to better recall on a final test than if they had only studied “suds-bath”. This effect has been demonstrated with a variety of materials including fictional questions with no correct answer (Kornell, et al., 2009), obscure English words and their meanings (Potts & Shanks, 2014), and foreign language words and their English translations (see for example, Seabrooke et al, 2022). The significant boosts in performance in recall tests show that pretesting may be a legitimate alternative to regular studying when it comes to improving memory of learned items. Exploring and understanding the processes behind why the pretesting effect takes place can lead to it being replicated in classroom and educational settings.

Although the pretesting benefit has been widely replicated, the mechanism behind this effect has not been established. Several theories have been proposed to explain this effect. One prominent theory is error correction theory. Error correction theory states that when a retrieval attempt takes place and is failed, the failed attempt leads to an error signal taking place, which enhances the learning attempt that follows (Kornell, et al., 2009).

Another theory known as additional-cue theory suggests that the word participants generate during the pretest can be used as an additional cue at test to help recall the target (Soraci, et al., 1999).

Another theory, search set theory, states that during the retrieval attempt participants activate a set of associated candidate items, one of which is the target, which are primed when they study the cue-target pair. That is, during a pretest, the retrieval attempt activates a set of candidate items that are semantically related to the cue. These items include both strong and weak associate words. Although a strongly related candidate is produced, both strong and weakly associated items are activated and when a weakly associated item is studied as the target, encoding is enhanced due to the semantic activation it had received during the pretest (Grimaldi & Karpicke, 2012).

Grimaldi and Karpicke (2012) conducted a series of experiments evaluating evidence for these theories. Specifically, they found that pretesting led to significantly improved performance on a final recall test compared to an only study condition when cue and target are semantically related, but pretesting did not improve recall when they were unrelated. This finding provides evidence against error-correction theory because performance did not change for unrelated words. In a second experiment, they tested the additional-cues theory by adding two new conditions. In one condition, the study lure condition, participants studied the cue paired with an incorrect lure (a strong associate to the cue) prior to being given the correct cue-target pair. In another condition, the constrained-pretest condition, participants were given the first two letters of a strongly associated lure with the cue during the pretest. On a final test, not only did pretesting outperform all other conditions (study, study lure, and constrained-pretest), and study lure and constrained-pretest, two conditions in which an additional word (lure) was present, led to worse performance on the final cued recall test. This finding was interpreted as evidence against both additional-cues theory and error-correction theory because giving participants an additional cue and leading participants to make an error both resulted in worse performance.

Another potential explanation for why pretesting leads to better recall, known as the elaborative generation hypothesis, is that retrieval (via pretesting) activates concepts related to the cue that become associated with the target. These items can be used during the test to help retrieve the target (Potts, et al., 2019). The elaborative generation hypothesis and the search set theory are similar in that, according to both theories, potential candidates (i.e., words that are semantically associated with the cue) are activated during a retrieval attempt. However, the elaborative generation hypothesis attributes the mnemonic benefit to associations made between the candidate items and the target, whereas the search set theory attributes it to better encoding of the recently activated target word during subsequent study.

Both the elaborative generation hypothesis and the search set theory assume that semantically associated items are activated in memory during the pretest. The current study aimed to test this assumption. Specifically, we reasoned that if pretesting activates weakly related target words, then those words should show priming during a lexical decision task (LDT). In a lexical decision task participants are given a string of letters on screen and asked to identify if the string of letters is a word or a nonword, as quickly as possible. This task can be used to measure semantic activation by comparing reaction time for primed words to unprimed baseline words (Meyer, & Schvaneveldt, 1971). If a word has been semantically activated prior to the lexical decision, then participants should be faster to identify it – this reaction time benefit is called "priming". If semantic activation theory is correct, then weakly related target words should show priming on a lexical decision task after a pretest. More specifically, we predicted that following semantic retrieval (i.e., an attempt at guessing the target word), both weak and strong semantic associates of the cue would be more quickly identified during a subsequent lexical decision task than words that are unrelated to the cue.

We also predicted that studying a cue-target pair would lead to priming of the target item on a lexical decision task because it had previously been presented (i.e., during the study trial). This is a priming effect called repetition priming (see for example, Scarborough, Cortese, & Scarborough, 1977). Further, we expected that when participants made a lexical judgement after both semantic retrieval and study of the cue-target pair, they would be faster to identify target words than unrelated words due to semantic priming and/or repetition priming.

Method

Participants

One-hundred twenty English-speaking Purdue University undergraduate students participated in exchange for partial credit towards an introductory psychology course. Subjects completed the experiment in-person on individual computers in groups of one to seven people.

Materials and Design

We created a list of 36 sets of items that were used as stimuli. Each set included a cue word (e.g., *suds*), a weakly related target word (e.g., *bath*), a strongly related candidate word (e.g., *soap*), an unrelated word (e.g., *leaf*), and a pronounceable nonword (e.g., *chaumb*). Items were selected from a normative word-association database (Deyne, et al, 2019) and nonword database (Rastle, Harrington, & Coltheart, 2002). Weakly related target words had an approximate forward associative strength of .05 (meaning that, in a free association task, they were produced in response to the cue by about 5% of participants). The strongly related candidate words had an average forward associative strength of .36 and were each the word with the highest forward associative strength for their corresponding cue. All unrelated words had a forward associative strength of zero to their corresponding cue.

We used a completely within-subject design such that, for each participant, a third of the items were presented in the study only condition, a third were presented in the pretest (retrieve then study) condition, and a third were presented in the retrieve only condition. For each of those three conditions, on the subsequent lexical decision task trial, a quarter of the time the stimulus was the weakly associated target word, a quarter of the time it was the strongly associated candidate word, a quarter of the time it was an unrelated word, and the remaining quarter of the time it was a nonword. The presentation of items within these conditions was counterbalanced across participants such that all sets appeared equally in each condition. The presentation order of items was randomized for all participants.

Procedure

We followed the general procedure described by Grimaldi and Karpicke (2012; see also Kornell et al., 2009) except that after each trial of the learning task, participants completed one trial of a lexical decision task (see Experiments 1a and 1b of Lehman & Karpicke, 2016 for a similar design).

At the beginning of the experiment participants were instructed that they would be alternating between two types of tasks. In one task they would be learning pairs of words and in the other task they would be making decisions about whether a string of letters formed a valid English word. They were told that during the learning task, they would sometimes be shown the pair of words to study (study only condition), other times they would be given one word and asked to guess what the other word might be. They were told that after making a guess they would sometimes be shown the correct pair to study (pretest condition) but other times they would not (retrieve only condition). After receiving instructions about both tasks, participants completed practice trials and were given a chance to reread the instructions and ask questions if necessary.

During the experiment, participants alternated between trials of the different learning task conditions and trials of the lexical decision task. For study only trials, participants saw a pair of words, consisting of a cue word and a weakly related target word, on screen for 5 seconds. They then saw a fixation cross on screen for 1 second, followed by a trial of the lexical decision task. For retrieve trials, participants saw a cue word on screen with a blank box under it. Participants were given 7 seconds to try to guess the correct target word and type their guess into the text box. If it was a retrieve only trial, they were shown a fixation cross followed by a lexical decision task trial, if it was a pretest trial, then the participant saw the correct cue-target pair for 5 seconds, before seeing a fixation cross, a lexical decision task trial.

During each trial of the lexical decision task, participants saw a string of letters and were asked to respond by pressing either “z” or “m” on the keyboard whether the letters made a valid English word (m) or not (z). On some trials participants saw a nonword. On the rest of the trials the word was either unrelated to the cue word for the previous learning task trial, the weakly associated target word, or the strongly related candidate word. After making a response, participants began a new trial of the learning task.

After completing 36 learning and lexical decision task trials, participants received a final cued recall test. Participants were shown the cue word for each of the 24 cue-target pairs that they had studied (either in the study condition or the pretest condition) and were asked to type the correct target word that it had been paired with. Each cue was presented one at a time in a new random order along with a text box for participants to make their response. Participants had 7 seconds to remember and type the target word.

Results

Final Cued-Recall

The scores from the final cued-recall test were subjected to a paired samples t-test. Consistent with previous experiments, we found that pretesting (retrieval followed by study) led to better performance than study alone, $t(119) = 10.45$, $p < .001$, $d = .95$ (Figure 1).

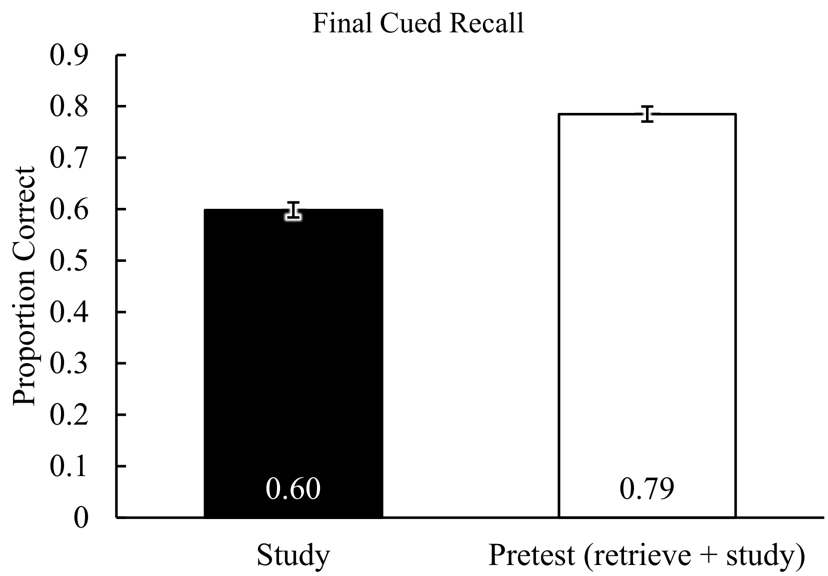

Figure 1: Proportion of words correctly on the final cued-recall test for the study and pretest conditions. Error bars represent standard error of the mean (SEM).

Lexical Decision Reaction Time

Reaction time (RT) to words in the lexical decision task were subjected to a 3x3 Repeated Measures ANOVA with learning tasks (study, pretest, and retrieve) and lexical decision word type (unrelated, target, and candidate) as within-subject factors (Figure 2).

We observed a main effect of the learning activity ($F(2,238) = 24.10$, $MSE = 69180.36$, $p < .001$, $\eta_p^2 = .17$) such that participants were faster to respond on lexical decision task trials following a learning trial in which they studied a cue-target pair than a trial in which they made a guess before studying a cue-target pair ($t(238) = -2.75$, $p = .007$) and participants were faster to respond on lexical decision task trials following a learning trial in which they made a guess before studying a cue-target pair than a trial in which they only made a guess but did not study the cue-target pair, $t(238) = -4.33$, $p < .001$.

There was no main effect of word type on RT in the lexical decision task, $F(2,238) < 1$, $MSE = 52412.09$, $p = .693$, $\eta_p^2 = .003$. In other words, there was no difference in RT on the lexical decision task for words that were unrelated to the cue compared to words that were weakly or strongly related to the cue.

These effect were qualified by an interaction between the type of word in the lexical decision task and the type of learning activity performed, $F(4,476) = 4.98$, $MSE = 50484.59$, $p < .001$, $\eta_p^2 = .04$. Specifically, after studying a cue-target pair, participants were faster in a lexical decision task at identifying words that were weakly related targets ($t(238) = 3.01$, $p = .003$) and strongly related candidates ($t(238) = 2.06$, $p = .041$) than words that were unrelated to the cue. In other words, we found priming of both target and candidate items following a study trial. The priming effect for target words compared to the unrelated baseline words was in line with our predictions that study of a cue-target pair would produce repetition priming.

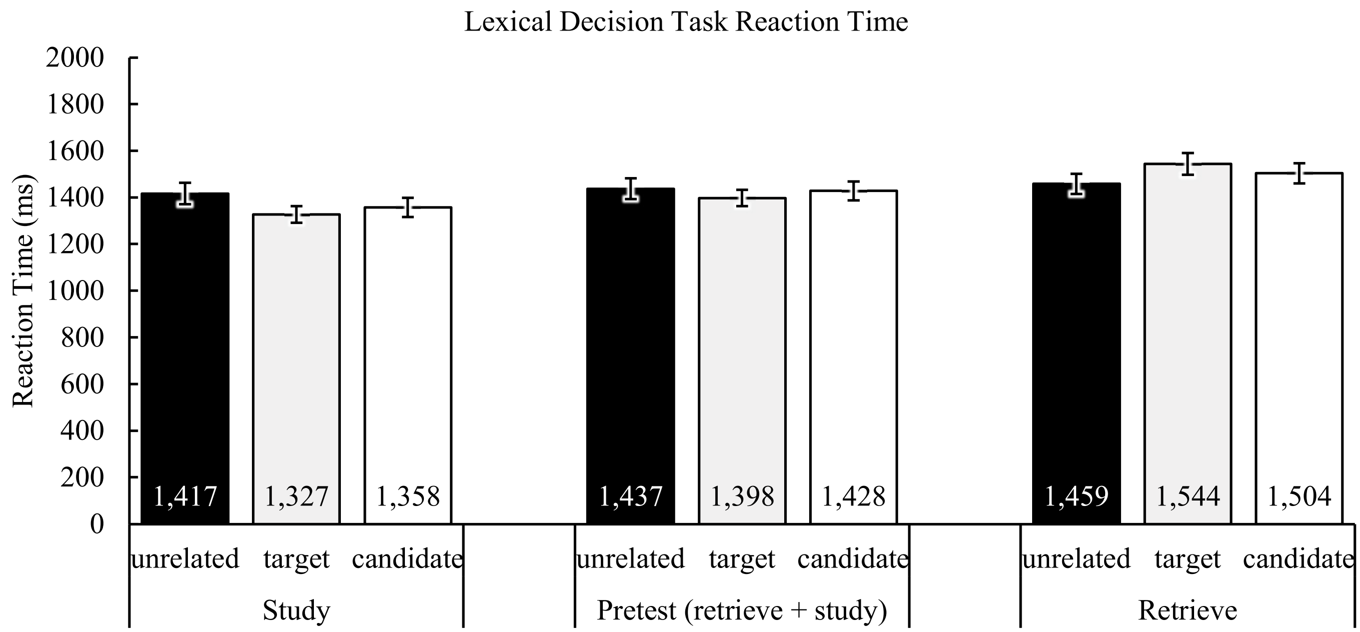

Figure 2: Reaction time on the Lexical Decision Task separated by learning task type of proceeding trial. Error bars show within-subjects standard error.

In contrast, and contrary to our prediction that target words would show priming after a pretest in which participants made a guess based on a cue word and then studied a cue-target pair, there were no significant differences in RT when identifying an unrelated word compared to either target words ($t(238) = 1.27$, $p = .207$), or candidate words, $t(238) < 1$, $p = .745$. We also did not observe a priming benefit for target words following a trial in which participants made a retrieval attempt along (without study).

Contrary to what was predicted based on semantic activation accounts, after making a guess based on a cue word (with no feedback about the correct target was), participants were slower to identify weakly related targets compared to unrelated words, $t(238) = -2.66$, $p = .009$. There was no difference in RT for unrelated words compared to candidate words, $t(238) = -1.59$, $p = .115$.

Discussion

Although we replicated the pretesting effect on a final cued recall test, we did not find any evidence to support the hypothesis that this pretesting benefit is due to semantic priming of associated word during a retrieval attempt.

As predicted for the study condition, the target words showed repetition priming when compared to the baseline control condition in the lexical decision task. However, participants were also faster to react to strongly related candidate words. For the pretesting condition, it was predicted that targets would be primed due to semantic and/or repetition priming. Instead, in the pretesting condition, neither the target nor the candidate items were reacted to faster compared to the baseline, unrelated, items.

Based on semantic association accounts of the pretesting effect, we predicted that both weakly related targets and strongly related candidates should be primed in the lexical decision task immediately following a retrieval attempt in response to the cue. We found no evidence of a priming benefit for semantically related words. In fact, we found negative priming for the target items compared to the baseline items. In other words, participants were slower to respond to the target after completing a retrieval attempt.

In short, although we observed a large mnemonic benefit on the final cued recall test for items that were presented compared to those that were only studied, there was no evidence of semantic activation (i.e., no priming on the LDT) of the target after a retrieval attempt (regardless of whether this attempt was followed by study of the cue-target pair). These results do not support the idea that targets receive partial activation during a pretest, leading to enhanced subsequent encoding during study. The findings of this study are strong evidence against semantic activation-based accounts of the pretesting effect. A new theory is necessary to be able to explain the results from this study, as no current theory supports these findings.

References

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, *41*(4), 977-990. https://doi.org/10.3758/BRM.41.4.977

De Deyne, S., Navarro, D.J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods* **51**, 987–1006 (2019). https://doi.org/10.3758/s13428-018-1115-7

Grimaldi, P., & Karpicke, J. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505-513. https://doi.org/10.3758/s13421-011-0174-0

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of experimental psychology. Learning, memory, and cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729

Lehman, M., & Karpicke, J. (2016). Elaborative retrieval: Do semantic mediators improve memory? *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *42*(10), 1573-1591. https://doi.org/10.1037/xlm0000267

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*(2), 227–234.

https://doi.org/10.1037/h0031564

Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect? Journal of Experimental Psychology: Learning, *Memory, and Cognition*, 45(6), 1023–1041. https://doi.org/10.1037/xlm0000637

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143*(2), 644–667. https://doi.org/10.1037/a0033194

Rastle, K., Harrington, J., & Coltheart, M. (2002) 358,534 nonwords: The ARC Nonword Database, *The Quarterly Journal of Experimental Psychology Section* A, 55:4, 1339 1362, https://doi.org/10.1080/02724980244000099

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3*(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Seabrooke, T., Mitchell, C.J., Wills, A.J. et al. The benefits of impossible tests: Assessing the role of error-correction in the pretesting effect. *Memory & Cognition* **50**, 296–311 (2022). https://doi.org/10.3758/s13421-021-01218-6

Soraci, S. A., Carlin, M. T., Chechile, R. A., Franks, J. J., Wills, T., & Watanabe, T. (1999), Encoding Variability and Cuing in Generative Processing, *Journal of Memory and Language*, *41*(4), 541-559, https://doi.org/10.1006/jmla.1999.2661.