Sorry! You lost me at restudy: The power of engagement during successive study

Alejandro Carranza (a3carranza@ucsd.edu)
Department of Psychology, University of California, San Diego

Emma H. Geller (egeller@ucsd.edu)
Department of Psychology, University of California, San Diego

Timothy Rickard (trickard@ucsd.edu)
Department of Psychology, University of California, San Diego

Abstract

The benefit of retrieval practice over restudy has been demonstrated across a variety of materials and settings. However, past research regarding the efficacy of *repeated* retrieval practice over *repeated* restudy has failed to consider participant engagement during passive restudy. Over four rounds, participants studied a list of 76 word-pairs using passive or engaged restudy (answering a semantic yes/no question about each pair). Participants who restudied with semantic engagement performed markedly better on a final cued-recall test than those who used passive restudy. Our findings suggest that the benefit of testing in the current literature may be due in large part to widespread use of an inefficient form of restudy.

Keywords: Testing effect; levels of processing; repeated restudy; cued recall; educational psychology

Introduction

The testing effect (or retrieval practice) is a robust psychological phenomenon that has been well documented for decades across a range of materials including word-pairs, triplets, and passages (Roediger & Karpicke, 2006; Rowland, 2014). In a typical testing paradigm, participants first encode (study) to-be-learned material (e.g., word-pairs), then restudy half of the pairs and are tested on the other half, and finally are given a final test to measure their recall performance. Researchers have demonstrated that testing oneself on material that has been previously studied leads to higher levels of retention and recall compared to simply restudying (Dunlosky et al., 2013; Karpicke & Aue, 2015).

Repeated Retrieval and Restudy

While much of the testing effect literature focuses on a single round of testing, researchers have also explored the benefits of repeated testing. Repeated testing, that is testing on the same material several times, leads to greater retention test accuracy than does either a single round of testing or repeated rounds of restudy (Karpicke & Roediger, 2009; Larsen et al., 2009; Roediger & Karpicke, 2006; Wang & Yang, 2023). Furthermore, repeated testing has been found to produce faster retrieval times compared to repeated restudy (Kubik et al., 2018; Racsmány et al., 2018; van den Broek et al., 2013).

In repeated testing trials, participants are engaged with the material in the sense that they are asked to actively recall the information they have learned. This engagement can be measured by response accuracy. However, in repeated trials that involve passive restudy (the nearly ubiquitous case in the testing effect literature), participants make no response and thus researchers have no way to measure how intently they engage with the material. Rarely do participants engage in any kind of additional task during restudy. For example, Kubik et al. (2018) had participants engage with the material during repeated restudy by having them vocalize the presented word-pair or enact it. However, they did not compare this method of restudy to the typical passive restudy that is found in most experiments.

During passive restudy, it is possible that participants simply stop paying attention, particularly in the context of item repetition, whereas those in the Test condition must engage with the presented material to generate correct responses. Alternatively, participants may attend to to-be-learned materials during passive restudy but use ineffective encoding strategies. In either case, the current literature may have assessed the benefits of testing against a particularly weak form of restudy.

The Present Study

The present study sought to answer two research questions. First, would active semantic engagement during restudy lead to better accuracy on a final test compared to passive restudy? Second, does active engagement during restudy have a positive effect on final test retrieval time?

To induce engagement during restudy trials, we built upon the *levels of processing* framework proposed by Craik and Lockhart (1972). This framework states that memory strength is influenced by the depth of mental processing during encoding, varying from shallow (surface features) to deep (semantic meaning), even when learning is incidental rather than intentional. In a series of experiments, Craik and Tulving (1975) demonstrated that rates of retention were higher when

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

participants were prompted to answer a deep (semantic) prompt about each item (e.g. "Is this a living thing?") compared to when they were asked a shallow question (e.g. "Is the third letter a vowel?"). In the current study, we aimed to determine whether active semantic engagement during restudy can enhance recall relative to passive restudy. Participants either repeatedly studied a list of word-pairs (passive) or repeatedly studied while also answering a yes/no question about each word-pair (engaged restudy) before taking a final cued recall test on all pairs. By having participants answer a question about each word-pair during repeated restudy, we can ensure that they are engaging with the material when it is presented (provided that their answers are sensible).

The levels of processing framework suggests that final test accuracy will be higher in the Engaged Restudy condition. That outcome is not assured, however, because participants in the Passive condition are given no instructions about how to study. If they spontaneously adopt an effective, semantically based study strategy, such as word-pair interactive imagery, they might conceivably perform as well or even better than participants in the engaged condition. This possibility throws further light on the unique properties of passive restudy. Across participant individual differences, passive restudy might yield a range of engagement levels, from primarily mind wandering to use of sophisticated encoding strategies.

We also measured retrieval times (RTs) to determine whether any observed accuracy gain due to engaged restudy is accompanied by shorter RTs. Although repeated testing is known to improve RT, the RT consequences of repeated restudy have rarely been investigated, and it appears to have not been investigated at all for engaged vs. passive restudy.

Method

Participants

Sixty-two undergraduate students participated in this study in exchange for course credit. Four participants were removed for taking more than 15 seconds, on average, to answer questions in the final test, one was removed for answering questions on average under 350 ms, and one was removed for performing near chance in the engagement condition. Thus, data from 58 participants were analyzed (M_{Age} = 20.1 years, Female = 79.3%).

Materials

Word-Pairs 76 word-pairs were drawn from the Free Association Norms database (Nelson et al.,1998). Each of the 152 words was between four to seven letters in length and ranged from one to two syllables. Word-pairs had a similar average forward (.027) and backward (.029) associative strength.

Engagement Questions We created four yes/no questions requiring semantic processing: "Could at least one of these [words] be found in a house?", "Is at least one of these manufactured by humans?", "Could at least one of these be given as a gift?", and "Is it possible to purchase one of these from a supermarket?". Each question had a similar proportion of yes and no answers across the 76 word-pairs in the study.

Design and Procedure

The experiment was coded using jsPsych, an open-source JavaScript framework for creating behavioral experiments (v7.3, de Leeuw et al., 2023). Participants completed this experiment online using any desktop browser except for Safari. The experiment could not be completed on any mobile device, including smart tablets. Participants were provided a link with which to access the experiment upon signing up.

The experiment involved two stages: the study stage and the testing stage. For the study stage, participants were randomly assigned to do either passive or engaged restudy. Before beginning, participants were informed that they would study the word-pairs for four rounds before moving onto the final portion of the experiment. They were not explicitly told they would take a final test. In the Passive condition, participants were presented with 76 word-pairs in random order one at a time for six seconds each (see Figure 1). They then repeated this procedure three more times for a total of four rounds of study. Between each round of study, participants were given a 10 second break.

In the Engaged Restudy condition, participants were presented with the same 76 word-pairs in random order, one at a time, for six seconds each, across four repetition rounds. Throughout each study round, one of the four engagement questions appeared at the top of the screen. For each word-pair presented, participants were asked to answer "yes" or "no" to that question by pressing a key (see Figure 1).

If participants failed to provide an answer, the next word-pair was displayed after six seconds. If participants provided an answer in less than six seconds, the word-pair remained on screen for six seconds. This was done to ensure that participants in both conditions had the same viewing time for each word-pair. Questions were randomized such that each round had a different question, the order of questions was randomized across participants, and all trials within a round had the same question displayed. Again, participants were given a 10 second break between rounds. Following the completion of the study stage, every participant began the testing stage.

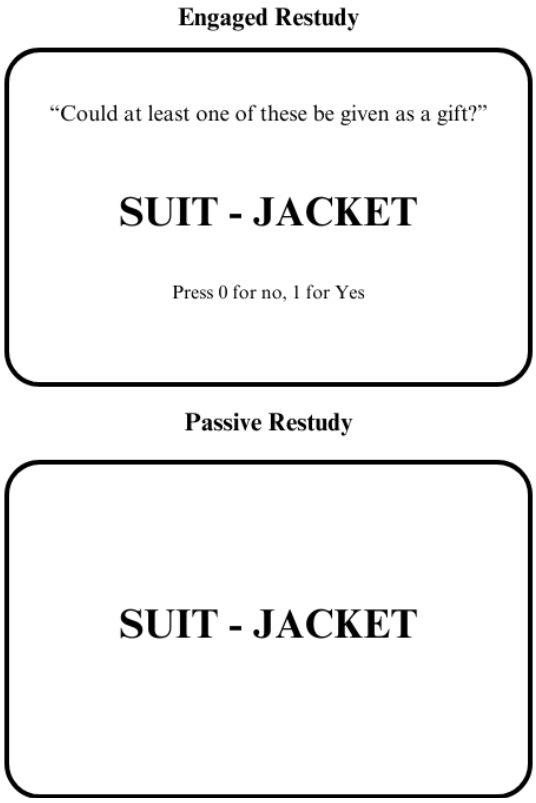

Figure 1: Engaged Restudy screen vs. Passive Restudy screen.

The testing stage began with a two-minute distractor task where participants were asked to solve a series of twenty-four simple subtraction and addition problems. Participants were then given a final cued-recall test for the 76 word-pairs. For the final test, participants were shown the left half of the word-pair (the cue) and asked to type out the right half (the target) and then press the enter key. A response was recorded as correct only if the spelling of the response exactly matched the spelling of the target word. RT was measured in milliseconds as the time from presentation of the cue word to pressing of the enter key. RT data was only analyzed for trials where the participant answered correctly. The order of word-pairs was randomized, and participants had unlimited time to answer. No feedback was provided.

Results

Engagement Check

Before analyzing the final data, we first examined the accuracy of participants' question responses in the Engaged Restudy condition of the study phase. Correct answers for each engagement question and word-pair combination were judged by the primary author.

To correct for guessing, each participant's *corrected score* was calculated as their observed proportion correct minus their observed proportion incorrect, yielding a corrected score of zero when exactly half of the responses were correct. This corrected score was then analyzed (See Figure 2). One participant scoring below 0.3 was excluded from further analyses due to suspected lack of cooperation. This outlying score is visible in the distribution shown in Figure 2. No participants had corrected scores below zero, indicating that all participants attempted to answer the engagement questions correctly.

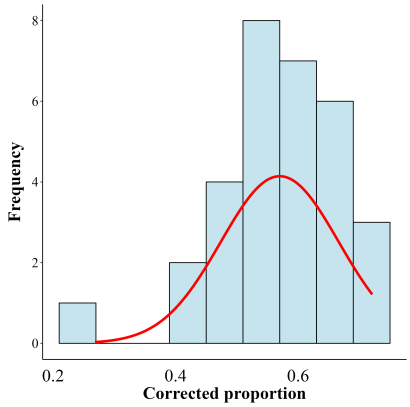

Figure 2: Histogram of corrected proportions for Engaged Restudy.

Final Test Accuracy

We used R (v4.1.3, R Core Team, 2023) and lme4 (Bates et al., 2015) to perform a binomial generalized linear mixed effects analysis of condition on final test proportion correct. In the null model, participants and items were random effects to account for variability in difficulty across word-pairs and individual differences in memory across participants. In the alternative model, we added condition (without interaction terms) as a fixed effect. The full model is shown in Table 1.

Table 1: Full Accuracy Model

Model Equation					
Accuracy ~ Condition + (1 | Participant) + (1 | Item)					
Fixed Effects					
	Est/Beta	*SE*	*95% CI*	*z*	*p*
Intercept	0.3153	0.27	-0.22, 0.88	1.17	0.24
Passive	-1.2620	0.36	-1.99, -0.53	-3.46	**<0.01**
Random Effects					
		Variance		*SD*	
Participants (Intercept)		0.5999		0.7745	
Items (Intercept)		1.8310		1.3531	

Participants in the Engaged Restudy condition were more accurate on their final test performance (M = 0.58, 95% CI [0.47, 0.70]) than participants in the Passive condition (M = 0.28, CI 95% [0.18, 0.40]) (see Figure 3). A likelihood ratio test indicated that the model that included condition provided a significantly better fit to the data than the model without it

(χ^2 (1) = 10.78, p < 0.01). The effect size for the difference between the conditions was d=0.96, which is considered to be a large effect.

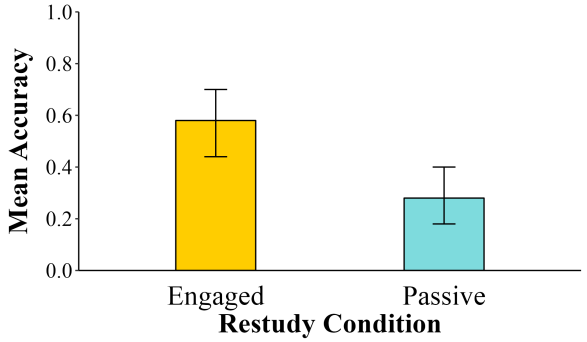

Figure 3: Mean Accuracy by condition. Error bars denote 95% confidence intervals.

Final Test Retrieval Time

We then performed a linear mixed effects analysis of condition on retrieval time using R (v4.1.3, R Core Team, 2023) and lme4 (Bates et al., 2015). We first logarithmically transformed the RT data to approximate the assumption of normality. In the null model, participants and items were entered as random effects. For the alternative model, condition was used as a fixed effect. Table 2 displays the full model.

Table 2: Full RT Model

Model Equation					
Log RT ~ Condition + (1 | Participant) + (1 | Item)					
Fixed Effects					
	Est/Beta	*SE*	*95% CI*	*t*	*Pr(>|t|)*
Intercept	3.47083	0.02	2671.38, 3273.93	156.32	<0.001
Passive	0.05227	0.03	0.98, 1.30	1.71	0.093
Random Effects					
			Variance	*SD*	
Participants (Intercept)			0.011507	0.10727	
Items (Intercept)			0.004504	0.06711	

On average, participants in the Engaged Restudy condition had faster retrieval times (M = 2951.21, CI 95% [2691.53, 3311.31]) than participants in the Passive condition (M = 3311.31, CI 95% [3019.95, 3715.35]) (see Figure 4). However, a likelihood ratio test indicated that the model including condition did *not* provide a significantly better fit for the data than a model without it (χ^2(1) = 2.86, p = 0.09). The effect size was d = 0.17, which is considered to be a small effect.

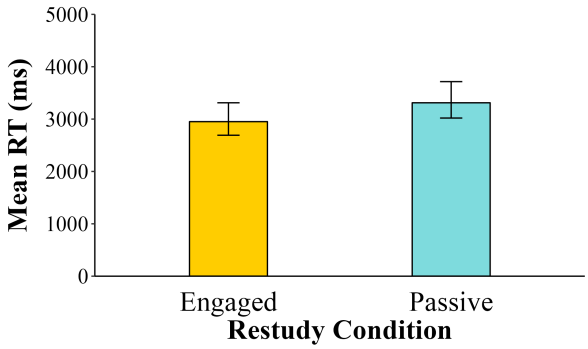

Figure 4: Mean RT by condition. Error bars denote 95% confidence intervals.

Discussion

We investigated whether engaged restudy yields more accurate and (or) faster test performance than does passive restudy. Participants studied 76 word-pairs by using either passive or engaged restudy before being given a final cued-recall test. In line with our predictions, we found that participants who were asked to repeatedly restudy with semantic engagement performed significantly better on a recall test than did those who used passive restudy. Additionally, participants in the Passive and Engaged Restudy conditions did not differ significantly in their retrieval times.

These findings have several theoretical implications. The better test performance of participants in the Engaged Restudy condition indicates that the existing literature on repeated testing may be overestimating the benefits of repeated testing relative to repeated restudy, at least for this type of cued recall task. Researchers may be failing to fully engage participants during restudy, resulting in attentional dropout – or mind wandering -- that negatively impacts final test performance. Higham et al. (2023) similarly found that restudy outperformed retrieval practice when participants were asked to provide a judgment of learning during successive relearning. This suggests that being asked to think about the material in a more profound manner during passive restudy, either through semantic questions or judgments of learning, results in better performance, in line with our own findings.

Passive study is likely used extensively under ecologically valid conditions, as for example when students study facts or term-definition pairs in preparation for an exam. It is thus a reasonable control condition in experiments with an applied orientation. However, our results indicate that study can be modified to induce greater levels of memory retention. If that is broadly the case, then cued recall testing may not be as superior a learning mechanism as the literature suggests. Our findings open the door to investigating novel approaches to restudy that may surpass testing, particularly in areas where testing may fall short, as in the case of transfer of test-based learning under some circumstances (Pan & Rickard, 2018).

An important note to make regarding the current study is that in neither condition were participants limited in the processes they could use to encode the word-pairs. During engaged restudy, participant attention was directed towards the word-pair on screen to ensure they engaged with the material at a semantic level. However, we did not provide them with specific instructions on how best to remember each pair. As such, our engagement manipulation kept their attention without restricting their encoding strategies. It is possible that even greater learning through restudy could be achieved using yes/no questions that direct participants to focus on more specific relations between cue and target elements, e.g., Do the two words interact in the natural environment (bird and tree)?

The non-significant difference in retrieval times between passive and engaged restudy is notable given the large advantage in proportion correct for the engaged condition. Based on these results, it may be that response time gains with repetition occur primarily through retrieval practice.

Future Directions

There are numerous avenues for further exploration. Most notably, future studies should investigate how repeated engaged restudy compares to repeated retrieval practice. It is possible that engaged restudy can rival retrieval practice on the final test. It's equally plausible that engaged restudy, while better than passive, will not match testing with feedback.

Second, this study focused on only one form of engagement, and we cannot determine whether the outcomes observed are due to the semantic engagement itself or rather simply to the requirement for participants to attend to the words pairs in some fashion, minimizing the possibility of a complete lack of engagement. In the former account, semantically engaged restudy should be much more effective than a shallower form of engagement within the levels of processing framework, such as questions at the lexical or graphemic level.

Third, the results of our study are only applicable to literature that focuses on repeated passive vs. repeated retrieval practice, as we did not have a condition with only one round of restudy. It's possible that engaged restudy is beneficial only in the context of repeated rounds of study. Future research should explore different ways in which to keep participants engaged during restudy.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. 10.18637/jss.v067.i01

Craik, F. I.M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I.M., & Turving, E. (1975). Depth of Processing and the Retention of Words in Episodic Memory. *Journal of Experimental Psychology: General*, *104*(3), 268-294. https://doi.org/10.1037/0096-3445.104.3.268

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software*, *8*(85), 5351. https://doi.org/10.21105/joss.05351.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Association For Psychological Science*, *14*(1), 4-58. 10.1177/1529100612453266

Higham, P. A., Fastrich, G. M., Potts, R., Murayama, K., Pickering, J. S., & Hadwin, J. A. (2023). Spaced Retrieval Practice: Can Restudying Trump Retrieval? *Educational Psychology Review*, *35*(98). https://doi.org/10.1007/s10648-023-09809-2

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

Karpicke, J. D., & Roediger, H. L. (2009). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966-968. 10.1126/science.1152408

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17-29. https://doi.org/10.1016/j.jml.2012.02.004

Kubik, V., Jonsson, F. U., Knopf, M., & Mack, W. (2018). The Direct Testing Effect Is Pervasive in Action Memory: Analyses of Recall Accuracy and Recall Speed. *Frontiers in Psychology*, *9*(1632). 10.3389/fpsyg.2018.01632

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomized controlled trial. *Medical Education*, *43*, 1174–1181. doi:10.1111/j.1365-2923.2009.03518.x

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/.

Racsmány, M., Szollosi, Á., & Bencze, D. (2018). Retrieval Practice Makes Procedure From Remembering: An Automatization Account of the Testing Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 157–166. http://dx.doi.org/10.1037/xlm0000423

R Core Team. (2023). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. https://www.R-project.org/

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rowland, C. (2014). The Effect of Testing Versus Restudy on Retention: A Meta-Analytic Review of the Testing Effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

van den Broek, G. S.E., Segers, E., Takashima, A., & Verhoeven, L. (2013). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, *22*(7), 803-812. http://dx.doi.org/10.1080/09658211.2013.831455

Wang, L., & Yang, J. (2023). The influence of repeated study and repeated testing on the testing effect and the transfer effect over time. *Memory and Cognition*. https://doi.org/10.3758/s13421-023-01477-5