Retrieval Practice and Context Reinstatement

By

Jason M. Ayala

* * * * * * * * *

Submitted in partial fulfillment

of the requirements for

Honors in the Department of Neuroscience

UNION COLLEGE

June 2023

ABSTRACT

AYALA, JASON Retrieval Practice and Context Reinstatement. Department of Neuroscience, June 2023.

ADVISOR: ZACHARY BUCHIN

Retrieval practice is a powerful learning strategy, but the mechanism(s) behind it are not fully understood. One account of this “testing effect” (i.e., enhanced memory for information that was previously retrieved vs. restudied) is the episodic-context account (Karpicke et al., 2014). According to this theory, successful retrieval requires contextual reinstatement which updates the target memory representation to include features of the current test context along with features of the initial study context. The resulting composite trace provides varied features that are more likely to match those on the final test. However, the specific features necessary for reinstatement remain unclear. The current study was designed to clarify the nature of these features using a three-phase paradigm. In phase 1, participants studied targets paired with different cues (phonemic vs. semantic) and presented in one of two different lists. In phase 2, participants either: (a) restudied the words; (b) restudied and made cue discrimination judgments; (c) restudied and made list discrimination judgments; or (d) freely-recalled the words. In phase 3, a final recognition test assessed item and context memory (i.e., cue type and list number). Unconditionalized analyses of final test performance found no differences between learning conditions. However, when only looking at words successfully retrieved at practice and thus subjected to reinstatement, free recall practice generally outperformed all other conditions on item and context memory, aside from the Restudy + Cue Discrimination condition. These results suggest that semantic, rather than temporal, features may be most relevant to the learning benefits of retrieval practice.

Retrieval Practice and Context Reinstatement. Department of Neuroscience, June 2023.

Retrieval practice is one of the most potent and effective ways to enhance learning. The testing effect refers to the widely documented finding that retrieving information leads to better performance on a future test vs. restudying that information (McDermott, 2021; Rowland, 2014). The benefits of retrieval practice have been known for over a century since the work of Abbot (1909) and Gates (1917), but the mechanism(s) behind those benefits are less understood.

Several theories have been proposed to explain the mechanisms behind the learning benefits of retrieval practice. For example, the elaborative retrieval account (Carpenter, 2009) proposes that retrieval attempts activate semantic associates of the memory cue, which can then be used to access the target. On a final test, recalling the target is easier because of these additional retrieval routes activated during practice. Because restudying does not involve a retrieval attempt, no similar benefits occur in the restudy condition. However, recent research has found evidence against this explanation by showing that retrieval is more beneficial than other similar elaborative strategies (Karpicke & Blunt, 2011) and that dividing attention during retrieval practice does not minimize the benefits seen on the final test, even though elaborative processing requires attentional resources and has been shown to be especially susceptible to divided attention and distraction (e.g., Buchin & Mulligan, 2017).

A second theoretical explanation of the testing effect implicates context reinstatement during retrieval practice as the key mechanism behind the benefit. Specifically, the episodic context account (Karpicke et al., 2014) states that in order to successfully retrieve the target information on the practice test, one must reinstate the context associated with that information.

After successful retrieval, the features associated with the practice test context are added to the item's reinstated memory trace. On a later final test, the item can be retrieved by reinstating either set of context features (i.e., those from initial study or those from the practice test), which aids recall by restricting the search set of potential targets. This leads to larger benefits from retrieval practice than restudying because, in the latter condition, no retrieval is necessary during practice and thus no context reinstatement occurs.

The episodic context account provides a comprehensive explanation for various findings in the retrieval practice literature. For example, the account offers insights into the role of retrieval mode in retrieval practice effects. Retrieval mode refers to the cognitive state in which individuals intentionally recall a specific time and place when an event occurred (Tulving, 1983). Karpicke and Zaromb (2010) conducted a series of experiments in which they manipulated certain conditions for participants, including whether they read (studied) target items, generated the target items, or retrieved the targets from a previous study session. In the latter condition, participants were prompted to actively recollect the initial study memory and complete the fragment with the word they had studied. The authors found that intentionally retrieving the targets produced greater performance on the final test compared to either restudying those targets or generating targets without recollecting the study episode. Therefore, reinstating the original episodic context during the practice phase enhanced subsequent retention.

The episodic context account can also explain the consistent observation is that spaced retrieval, where practice is distributed over time, leads to better retention compared to massed retrieval, where practice is crammed together (Roediger & Karpicke, 2011). According to the context account, during spaced repetition, the temporal context undergoes more change, necessitating a higher degree of context reinstatement during retrieval. Additionally, spaced retrieval results in updated and more distinctive context representations compared to massed retrieval. The context theory also elucidates the positive effects of "effortful" initial retrieval tasks. For instance, free recall tests yield larger retrieval practice effects than recognition tests, weakly associated cues yield greater effects than strong associates, and initial recall with only the first letter of a target as a cue produces larger effects than recall with three letters of the target (Carpenter, 2009; Carpenter & DeLosh, 2006; Pan & Rickard, 2018). In all these cases, the conditions that produce larger retrieval practice effects involve learners engaging in a higher level of context reinstatement during the initial retrieval process.

Multiple prior studies have found support for the predictions stemming from the episodic context account (for review, see Karpicke, 2017). However, certain aspects of the theory require further exploration. For example, the word “context” is very broad and the specific features necessary for context reinstatement are unclear. Karpicke et al. (2014) focus on the “temporal context” associated with the target information, which can contain any contextual features bound to that specific temporal episode (i.e., the memory trace formed when initially restudying the target). The more time between initial study and later reinstatement, the more diverse the newly-added features will be due to the larger change in temporal context. On the final test, having these multiple different features aids retrieval more than having multiple similar features.

The challenge in validating this temporal explanation arises when trying to infer the specific features that were reinstated at practice. For example, several studies have varied the contextual features associated with the target information at initial study before having participants engage in subsequent retrieval practice or restudy (e.g., Akan et al., 2018; Whiffen & Karpicke, 2017). On a final recognition test, participants are asked about these context features after successfully recognizing a target item. If the retrieval practice condition remembers more context features, it is assumed that those features were reinstated at practice. For example, Akan et al. (2018) had participants initially study targets in different spatial locations on a computer screen before either restudying or retrieving those items at practice. They found better memory for spatial context in the retrieval practice group than the restudy group and concluded that this meant those features were used for reinstatement at practice and are important to the memory benefit from retrieval practice (Akan et al., 2018). Whiffen and Karpicke (2017) found a similar pattern of results when manipulating the initial list in which each item was studied.

Other similar paradigms have produced conflicting results. Hong et al. (2019) varied the initial font color of target items at initial study and had participants engage in subsequent restudy or retrieval practice. On the final test, there was no difference in context recall (i.e., the original font color) between the two learning conditions, leading Hong et al. (2019) to conclude that those features must not have been reinstated during retrieval practice. Brewer et al. (2010) obtained a similar finding by varying the gender of the voice used to provide the target items at initial study. Although these results seem to indicate that those features (i.e., font color and voice gender) were not used for reinstatement at retrieval practice, this assumption is potentially problematic. Even if those features were reinstated at practice, participants may not remember them on the final test.

To help address this ambiguity, the present study compared a typical retrieval practice condition (i.e., free recall) to different types of restudy + context feature discrimination conditions to provide indirect evidence of the features most critical to reinstatement at retrieval practice. Participants studied two lists of 20 cue-target word pairs for a later recognition test on the target words. In the subsequent practice phase, participants either engaged in restudy, restudy + list discrimination, restudy + cue discrimination, or retrieval practice. The final recognition test assessed item memory as well as context memory (i.e., the initial list the word was studied in as well as the initial cue type it was paired with). To infer which features were critical to reinstatement and thus contribute to the testing effect, final item and context memory in the two restudy + discrimination conditions were compared to the Retrieval Practice condition.

Method

Participants

Sixty-one participants were recruited from Union College and participated in exchange for either course credit or monetary compensation. Participants were about 20 years old (*M* = 19.77, *SD* = 1.13) and 54% identified as female (43% as male and 3% as non-binary). The racial and ethnic composition of the participants was as follows: White (77%), African-American (7%), Asian (7%), mixed (5%), Latino/a (3%), and Ashkenazi Jewish (2%).

Design and Materials

The key independent variable (i.e., learning condition during the practice phase) was manipulated between-subjects and had four levels: (a) Restudy; (b) Restudy + List Discrimination; (c) Restudy + Cue Discrimination; and (d) Retrieval Practice. All conditions had 16 participants except for the Restudy condition, which had 13 due to recruitment limitations.

Most materials were drawn from Su et al. (2020). Forty critical target words were selected. The average length, HAL frequency, and concreteness for target words was *M* = 4.35 (*SD* = 0.98, *Range* = 3 to 6), *M* = 69,607.83 (*SD* = 92,348.00, *Range* = 3,160 to 393,090), and *M* = 3.87 (*SD* = 1.14, *Range* = 2.93 to 6.92), respectively. For each target word (e.g., dinner), we selected a semantic cue (e.g., supper) and a phonemic (rhyme) cue (e.g., winner). The average length, HAL frequency, and concreteness for semantic cues was *M* = 5.13 (*SD* = 1.14, *Range* = 3 to 8), *M* = 44,902.30 (*SD* = 169,824.30, *Range* = 286 to 1,076,711), and *M* = 3.80 (*SD* = 1.04, *Range* = 1.54 to 5), respectively. The average length, HAL frequency, and concreteness for phonemic cues was *M* = 4.43 (*SD* = 0.90, *Range* = 3 to 6), *M* = 44,867.90 (*SD* = 105,544.47, *Range* = 1,363 to 580,704), and *M* = 3.70 (*SD* = 1.04, *Range* = 1.33 to 5), respectively. The average forward strength of semantic cue–target pairs was *M* = 0.56 (*SD* = 0.23, *Range* = 0.03 to 0.82). Forty foil words were intermixed with the target words on the final test. The average length, HAL frequency, and concreteness for foils was *M* = 4.90 (*SD* = 1.03, *Range* = 3 to 6), *M* = 210.78 (*SD* = 360.74, *Range* = 3 to 1,599), and *M* = 3.95 (*SD* = 1.15, *Range* = 1.62 to 5), respectively. Participants studied two 20-word lists in the Initial Study Phase that each contained 20 cue-target word pairs (10 semantic and 10 phonemic cues). List number and cue type were counterbalanced across conditions, such that each word was presented in each list and with each cue type an equal number of times.

Procedure

The experiment had three phases: (1) Initial Study Phase; (2) Practice Phase; and (3) Test Phase (see Figure 1). First, participants were given instructions about the upcoming recognition memory test that would assess their memory of 40 total target words that they would learn alongside cues via multiple rounds of studying. Participants studied each list of 20 cue-target word pairs a total of three times each. Each word pair was shown one at a time for 4 s. To ensure participants were paying attention, they were instructed to type each target word into a textbox as they studied. For example, the cue-target pair “near-hear” would be shown and participants would need to type “hear” in the box. Participants started with List 1, studying each pair one at a time for 4 s each. This process was repeated two more times for List 1, with the word pairs restudied in a newly randomized order, before moving on to List 2, which followed the same procedure as List 1. When finished, participants completed a math distractor task for 2 min. The long distractor interval was necessary to make the temporal context during retrieval practice different from the context at initial study.

Participants then moved on to the Practice Phase, where they engaged in their randomly assigned learning condition (i.e., Restudy, Restudy + List Discrimination, Restudy + Cue Discrimination, or Retrieval Practice). In the Restudy condition, participants were shown each of the 40 target words (without their prior cues) for 6 s of additional studying. As they restudied each target, they were instructed to type it into a textbox. The other two restudy conditions were the same but required participants to indicate either the prior list number (i.e., List 1 vs. List 2, for the List Discrimination condition) or cue type (i.e., Phonemic/Rhyme vs. Semantic/Meaning cue, for the Cue Discrimination condition) associated with each target in the Initial Study Phase. In the Retrieval Practice condition, participants had a total of 4 min to freely-recall and type each of the 40 target words they could remember from the Initial Study Phase. No feedback was given in any condition. This was followed by an additional 2 min of math problems before moving on to the final recognition test.

In the Test Phase, participants completed a self-paced, source recognition test that contained 80 total words (40 targets and 40 lures) presented in a randomized order. For each word, participants were asked if they recognized it from the prior phases and, if they indicated "yes", they were then asked about its context features (i.e., the list it was presented in and the cue type it was presented with during the Initial Study Phase). When finished, participants were thanked, debriefed, and compensated for their participation.

Results

All results were analyzed using the statistical software program JASP (JASP Team, 2023). Final performance on the final recognition test was analyzed in two ways: (1) using the typical unconditionalized scores (i.e., performance in the retrieval practice condition assessed all the words they initially studied, even if they did not retrieve them in Phase 2) and (2) using conditionalized scores (i.e., performance in the retrieval practice condition assessed only those words that were retrieved at practice in Phase 2). Although the latter measure is biased towards the Retrieval Practice condition (e.g., a smaller denominator and possible item-selection effects where only the easiest items are recalled in Phase 2 and then again in Phase 3), it also provides the most accurate assessment of the effects of context reinstatement during retrieval practice since only those words actually retrieved would have their context reinstated[1].

Unconditionalized Performance: Item Memory

Item memory on the final recognition test (see Table 1) was operationalized in two ways: (1) corrected hit rates (*Hit Rate - FA Rate*) and (2) sensitivity (*z[Hit Rate] - z[FA Rate]*). First, corrected hit rates were analyzed using a one-way ANOVA with Learning Condition (Restudy vs. Restudy + Cue Discrimination vs. Restudy + List Discrimination vs. Retrieval Practice) manipulated between-subjects. There was no significant main effect, $F(3, 57) = 1.37$, $p = .26$, indicating no difference in final corrected hit rates between any conditions. A similar pattern of results was found for sensitivity – no significant main effect of Learning Condition, $F(3, 57) = 1.77$, $p = .16$.

Unconditionalized Performance: Context Memory

Context memory on the final recognition test (see Table 1) was assessed separately for each context feature. First, proportion correct on final list discrimination was analyzed using a one-way ANOVA with Learning Condition (Restudy vs. Restudy + Cue Discrimination vs.

[1]In Phase 2, retrieval practice proportion correct was fairly low overall ($M = .23$, $SD = .09$), indicating that just a quarter of the words were successfully reinstated. For the two restudy + discrimination task conditions, average proportion correct was .63 ($SD = .15$) and .73 ($SD = .11$) for the list and cue discrimination tasks, respectively.

Restudy + List Discrimination vs. Retrieval Practice) manipulated between-subjects. There was no significant main effect, $F(3, 57) = 1.27$, $p = .29$, with similar performance across conditions. A similar pattern emerged for proportion correct on final cue discrimination – no significant main effect of Learning Condition, $F(3, 57) = 1.42$, $p = .25$. Taken together, there was no difference in final context memory between any of the four learning conditions.

Conditionalized Performance: Item Memory

The same analyses reported above were conducted again using the conditionalized scores (see Table 1), which only impacts the Retrieval Practice condition (e.g., if a participant only recalled 10 words during practice, their final test performance would only be focused on those specific 10 words since those were subject to the encoding effects of retrieval). First, conditionalized corrected hit rates were analyzed using a one-way ANOVA with Learning Condition manipulated between-subjects. There was a significant main effect of Learning Condition, $F(3, 57) = 3.89$, $MSE = .03$, $p = .013$, $\eta^2 = .17$. Post-hoc tests indicated that this was driven by greater performance in the Retrieval Practice condition compared to both the Restudy condition, $t = 2.89$, $SE = 0.07$, $p_{tukey} = .027$, $d = 1.08$, and the Restudy + List Discrimination condition, $t = 2.80$, $SE = 0.06$, $p_{tukey} = .034$, $d = 0.99$. No other comparisons were significant (i.e., all *t*s < 1.58, *p*s > .31). This suggests that, for the words subjected to context reinstatement, retrieval practice produced better item memory than both the Restudy and Restudy + List Discrimination conditions, but did not differ from the Restudy + Cue Discrimination condition.

An identical pattern emerged when analyzing conditionalized sensitivity scores. There was a significant main effect of Learning Condition, $F(3, 57) = 6.92$, $MSE = 1.69$, $p < .001$, $\eta^2 = .27$, and post-hoc tests indicated that this was due to greater performance in the Retrieval Practice condition compared to both the Restudy condition, $t = 4.05$, $SE = 0.49$, $p_{tukey} <.001$, $d =$

1.08, and the Restudy + List Discrimination condition, $t = 3.74$, $SE = 0.46$, $p_{tukey} = .002$, $d = 1.32$. No other comparisons were significant (i.e., all *t*s < 2.29, *p*s > .11).

Conditionalized Performance: Context Memory

Next, we re-analyzed final context memory using the conditionalized scores (see Table 1). For final list discrimination accuracy, a one-way ANOVA with Learning Condition manipulated between-subjects revealed a significant main effect of Learning Condition, $F(3, 57) = 8.40$, $MSE = .03$, $p < .001$, $\eta^2 = .31$. Post-hoc tests indicated that this was due to greater list discrimination performance in the Retrieval Practice condition than any of the restudy conditions. Specifically, list discrimination accuracy for the words reexperienced in phase 2 was greater in the Retrieval Practice condition than: (a) the Restudy condition, $t = 4.69$, $SE = 0.06$, $p_{tukey} < .001$, $d = 1.75$; (b) the Restudy + List Discrimination condition, $t = 3.75$, $SE = 0.06$, $p_{tukey} = .002$, $d = 1.32$; and (c) the Restudy + Cue Discrimination condition, $t = 3.14$, $SE = 0.06$, $p_{tukey} = .014$, $d = 1.11$. No other comparisons were significant (i.e., all *t*s < 0.60, *p*s > .33).

On the other hand, a slightly different pattern emerged for cue context features. As before, there was a significant main effect of Learning Condition, $F(3, 57) = 4.42$, $MSE = .03$, $p = .007$, $\eta^2 = .19$, but post-hoc tests indicated that the Retrieval Practice condition now only outperformed the Restudy condition, $t = 3.15$, $SE = 0.07$, $p_{tukey} = .014$, $d = 1.18$, and the Restudy + List Discrimination condition, $t = 2.95$, $SE = 0.06$, $p_{tukey} = .023$, $d = 1.04$. However, final cue accuracy did not differ between the Retrieval Practice condition and the Restudy + Cue Discrimination condition (nor between any other conditions, i.e., all *t*s < 1.58, *p*s > .26).

Discussion

The objective of this study was to assess the fundamental assumptions of the episodic context account (Karpicke et al., 2014). According to this account, when individuals engage in retrieval, they make an effort to reinstate the context from a previous learning episode. Successful retrieval leads to the updating of context representations, incorporating both features from the initial study context and the current practice test context. As a result, when individuals attempt to retrieve the same items in the future, the updated context representations enhance the retrieval of those items, leading to improved memory performance compared to situations where reinstatement at practice was not necessary (e.g., Whiffen & Karpicke, 2017). The present study provides results related to the nature of the context features critical to this reinstatement benefit.

First, when looking at the unconditionalized item memory scores, there were no differences between learning conditions. This is surprising, given the typical testing effect found with free recall practice (Carpenter & DeLosh, 2006) as well as the larger learning benefits from list discrimination practice than restudy found by Whiffen and Karpicke (2017). Why the discrepancy? One potential reason might be the nature of the final test. Since the present study assessed final context feature memory, we used a final source recognition test, whereas Whiffen and Karpicke (2017) used a final free recall test. Because a recognition test is typically easier than a recall test, it may have obscured the benefits from free recall practice by artificially enhancing performance across all conditions. Using a more difficult final test may have allowed those larger learning benefits to be seen on the final test. Further, proportion correct on the free recall practice test was generally low (~25%), especially compared to the 69% achieved by a similar testing effect study (i.e., Carpenter & DeLosh, 2006). This may have arisen due to the length of the word lists used in each study: 20 words per list vs. eight words per list. The longer lists in the present study may have also reduced list discrimination performance at practice. We observed lower list discrimination accuracy on the practice test (i.e., M = .63) than Whiffen and Karpicke (2017; i.e., M = .86). Having fewer words per list equates to having less to remember for the list discrimination task, likely contributing to the higher performance. This could explain the lack of an item memory benefit from list discrimination.

However, the analyses discussed above assessed performance on all initially studied words, regardless of if they were successfully retrieved on the practice test or not. This may be problematic if we want to assess the effects of, and on, context reinstatement, which only occurred for those words successfully recalled at practice. Therefore, to supplement the standard analyses, we conducted conditionalized analyses including only those words successfully recalled at practice.

When examining the conditionalized scores, the Retrieval Practice condition had better item memory than both the Restudy and Restudy + List Discrimination conditions. Thus, when only looking at the words successfully recalled on the free recall practice test, a testing effect emerged. Interestingly, there was no difference in item memory between the Retrieval Practice and Restudy + Cue Discrimination conditions, a point we return to later in the discussion.

Final context memory also did not differ between any conditions when assessing unconditionalized performance. Even though participants in the two discrimination conditions thought back to the original list or cue, it did not benefit their final context memory more than typical restudy or free recall practice. Differences only emerged in the conditionalized analyses, with the Retrieval Practice condition outperforming all other conditions in terms of final list memory, and all but the Restudy + Cue Discrimination condition on final cue memory. Thus, like with the item memory results, context reinstatement at practice seems to impact item and context memory in a way similar to thinking back to the original cue impacts memory.

This may mean that non-temporal (i.e., semantic) cues were reinstated at practice and are thus what aided performance on the final test. This finding is at odds with a central feature of the context account – the importance of temporal context. It is instead the semantic features here that may be more critical to context reinstatement and/or the benefits of reinstatement. Interestingly, there is a particular circumstance where the influence of temporal contiguity diminishes significantly: when semantic context is present (e.g., Hong, 2021). For example, when participants study a list comprising examples from various semantic categories, their free recall tends to display reduced reliance on temporal order (e.g., Healey & Uitvlugt, 2019; Polyn et al., 2011). Instead, participants tend to organize the recalled information based on semantic categories. This finding aligns with Hong's (2021) hypothesis that context reinstatement arises from the memory search process. It suggests that the type of reinstated context may depend on the contextual features employed to guide the recall process. Hong (2021) thus reasonably predicts that recalling items from a study list with strong semantic associations would lead to semantic contiguity being prioritized over temporal contiguity. Perhaps having participants study the targets alongside their cue words three times in a row made them engage in semantic processing and thus needed to rely less on temporal processing.

Limitations and Future Directions

Although the present study provided a novel indirect assessment of the context features critical to reinstatement by comparing multiple discrimination and recall conditions within a single experiment, it is not without its limitations. First, our sample may have been too small to detect effects of list discrimination and retrieval practice. For example, the average testing effect size is d = 0.50 (e.g., Rowland, 2014), and Whiffen and Karpicke (2017) found an effect of about d = 0.60 when comparing restudy to list discrimination. Power analyses indicate that at least 30 participants per between-subjects condition would be needed to detect effects of that size with power = .80 and α = .05. Perhaps our sample size of 16 per between-subjects condition was simply too small to detect these effects. Second, despite using multiple rounds of initial study, the low free recall practice test accuracy is problematic and prevents unambiguous conclusions regarding unconditionalized performance. Future research can attempt to boost initial recall performance via other methods, such as by using shorter word lists.

Implications

It is possible that other theories of the testing need to be further explored, as there are limitations to the context account. The primary and convergent retrieval model, proposed by Hopper and Huber (2019), suggests that the benefits of retrieval practice stem from enhanced primary retrieval rather than convergent retrieval. Primary retrieval involves associations between context and item features, while convergent retrieval involves associations between different item features. This model provides a different perspective from the episodic context account and highlights the importance of the learning mechanisms involved in memory recall. Another possibility is the cue-binding hypothesis, expanding on Hopper and Huber's (2019) model. According to this hypothesis, the benefits of retrieval practice do not solely rely on reinstating the episodic context but rather on binding the retrieved item to recall cues, regardless of the cue type. A recall cue is defined as any information that effectively guides the search process and leads to the correct target memory (Hong, 2021). With repeated retrieval, the association between the recall cue(s) and the retrieved item strengthens, resulting in faster recall. The recall cue can be related to the item's temporal, semantic, and spatial context or could be unrelated to the item's inherent characteristics but formed during retrieval, such as a self-generated thought (Zhang & Tullis, 2021).

The present study adds to the ongoing research on memory retrieval and emphasizes the importance of examining the potentially disparate effects of temporal, semantic, and other context features on later retrieval. Although we found that semantic features may be critical, we suggest that during retrieval, only certain aspects of the episodic context are reinstated, while others may not be. Other features that can be linked cohesively with other cues to activate the target memory could serve as an effective recall cue. The emphasis on investigating the role of temporal or semantic context in memory search and retrieval in existing research may be due to the prevalence of these context features in lab-based experiments, making them commonly examined and considered critical factors for recall. It underscores the necessity for future studies to delve deeper into understanding how these contextual factors affect our ability to accurately remember and retrieve information.

Tables

Table 1.

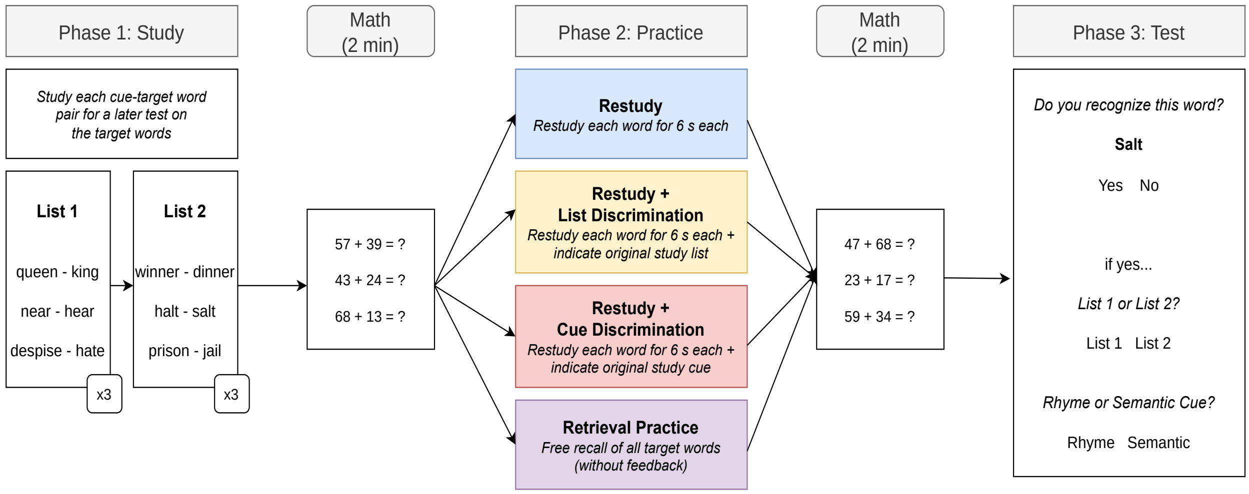
Final Recognition Test Performance: M (SD).

	Corrected Hit Rate (H - FA)	Sensitivity	List Proportion Correct	Cue Proportion Correct
Restudy	.73 (.20)	2.83 (1.01)	.46 (.18)	.57 (.19)
Restudy + List Discrimination	.75 (.24)	3.06 (1.42)	.53 (.14)	.59 (.19)
Restudy + Cue Discrimination	.85 (.14)	3.73 (1.43)	.56 (.16)	.69 (.17)
Retrieval Practice	.74 (.13)	2.83 (1.01)	.53 (.11)	.61 (.15)
Retrieval Practice (Conditionalized)	.92 (.10)	4.78 (1.18)	.74 (.17)	.78 (.17)

Figures

Figure 1.

Schematic Depiction of the Experimental Design.



References

Abbot, E. E. (1909). On the analysis of the memory consciousness in orthography. *The Psychological Review: Monograph Supplements*, *11*(1), 127-158. https://doi.org/10.1037/h0093013

Akan, M., Stanley, S. E., & Benjamin, A. S. (2018). Testing enhances memory for context. *Journal of Memory and Language*, *103*, 19-27. https://doi.org/10.1016/j.jml.2018.07.003

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory*, *18*(4), 385-393. https://doi.org/10.1080/09658211003702163

Buchin, Z. L., & Mulligan, N. W. (2017). The testing effect under divided attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1934–1947. https://doi.org/10.1037/xlm0000427

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563-1569. https://doi.org/10.1037/a0017021

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268-276. https://doi.org/10.3758/BF03193405

Gates, A. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, *(40)*, 104.

Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory and Cognition*, *47*(4), 719-737. https://doi.org/10.3758/s13421-019-00895-8

Hong, C. M. K. (2021). *Examining the Episodic Context Account of the Testing Effect: Does episodic recall enhance memory for context?* (Doctoral dissertation, Vanderbilt University).

Hong, M. K., Polyn, S. M., & Fazio, L. K. (2019). Examining the episodic context account: Does retrieval practice enhance memory for context? *Cognitive Research*: *Principles and Implications*, *4*(1), 1-9. https://doi.org/10.1186/s41235-019-0202-3

Hopper, W. J., & Huber, D. E. (2019). Testing the primary and convergent retrieval model of recall: Recall practice produces faster recall success but also faster recall failure. *Memory and Cognition*, *47*(4), 816-841. https://doi.org/10.3758/s13421-019-00903-x

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. Wixted (Ed.), *Cognitive Psychology of Memory, Vol. 2 of Learning and memory: A Comprehensive Reference* (J. H. Byrne, Series Ed.). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331(*6018), 772-775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 61, pp. 237–284). Elsevier Academic Press. https://doi.org/10.1037/xlm0000012

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language*, *62*(3), 227-239. https://doi.org/10.1016/j.jml.2009.11.010

McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology*, *72*, 609-633. https://doi.org/10.1146/annurev-psych-010419-051019

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*(7), 710–756. https://doi.org/10.1037/bul0000151

Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 766-775. https://doi.org/10.1037/a0022475

Roediger III, H. L., & Karpicke, J. D. (2011). Intricacies of spaced retrieval: A resolution. In *Successful Remembering and Successful Forgetting* (pp. 41-66). Psychology Press.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432-1463. https://doi.org/10.1037/a0037559

Su, N., Buchin, Z. L., & Mulligan, N. W. (2021). Levels of retrieval and the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(4), 652-670. https://doi.org/10.1037/xlm0000962

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *43*(7), 1036–1046. https://doi.org/10.1037/xlm0000379

Zhang, D., & Tullis, J. G. (2021). Personal reminders: Self-generated reminders boost memory more than normatively related ones. *Memory and Cognition*, *49*(4), 645-659. https://doi.org/10.3758/s13421-020-01120-7