ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pmem20

The role of mediators for the pretesting effect

Johannes Bartl, Oliver Kliegl & Karl-Heinz T. Bäuml

To cite this article: Johannes Bartl, Oliver Kliegl & Karl-Heinz T. Bäuml (01 Mar 2024): The role of mediators for the pretesting effect, Memory, DOI: 10.1080/09658211.2024.2323930

To link to this article: https://doi.org/10.1080/09658211.2024.2323930

Published online: 01 Mar 2024.

MEMORY
https://doi.org/10.1080/09658211.2024.2323930

Routledge
Taylor & Francis Group

Check for updates

The role of mediators for the pretesting effect

Johannes Bartl, Oliver Kliegl and Karl-Heinz T. Bäuml

Department of Experimental Psychology, Regensburg University, Regensburg, Germany

ABSTRACT

Taking a pretest (e.g., smoke – ?) before material is studied (smoke – *fog*) can improve later recall of that material, compared to material which was initially only studied. The goal of the present study was to evaluate for this pretesting effect the potential role of semantic mediators, i.e., of unstudied information that is semantically related to the study material. In all three experiments, subjects studied weakly associated word pairs (e.g., smoke – *fog*), half of which received a pretest. Subjects then either completed a recognition test (Experiment 1) or a cued-recall test (Experiments 2 and 3), during which they were presented with both the original study material and never-before-seen semantic mediators that were strongly related to the cue item of a pair (e.g., cigarette). Strikingly, presenting semantic mediators as lures led to higher false alarm rates for mediators following initial pretesting than study only (Experiment 1), and presenting semantic mediators as retrieval cues led to better recall of target items following pretesting than study only (Experiments 2 and 3). We argue that these findings support the elaboration account of the pretesting effect but are difficult to reconcile with other prominent accounts of the effect.

ARTICLE HISTORY
Received 27 September 2023
Accepted 14 February 2024

KEYWORDS
Testing effect; pretesting effect; elaboration; mediator

Tests can enhance memory. A particularly striking demonstration of the power of testing is the testing effect, which refers to the finding that tests of previously studied information can induce greater retention of the information on a later final test than does passive restudy of the information (e.g., Roediger & Karpicke, 2006). Over the last 20 years, the testing effect – which has sometimes also been referred to as posttesting effect – has been found to be a robust and general phenomenon, arising across a wide variety of study materials, including unrelated items, weakly associated word pairs, prose passages, and videos, and arising across a range of final test formats, such as free-recall tests, cued-recall tests, and recognition tests (for reviews, see Karpicke, 2017; Rowland, 2014).

Many explanations of the posttesting effect have been proposed over the years (for a review, see Karpicke, 2017), with the elaboration account ranking among the most prominent (e.g., Carpenter, 2011). The account assumes that during the initial test, the retrieval of a previously studied target item from a cue item (e.g., smoke – ?) is more probable to lead to the activation of information related to the cue item (e.g., *water*, *opaque*, *white*), and to become integrated with the originally studied cue-target pair (e.g., smoke – *fog*). When on a later final test, subjects are asked to produce the target item from the cue item, the additional information may become reactivated and potentially serve as an additional implicit cue, or semantic mediator, to access the target item, thus creating the posttesting effect.

Carpenter (2011) put the elaboration account to a test by examining how participants respond to the presence of semantic mediators during a final recognition test (Experiment 1) and a final cued-recall test (Experiment 2). In both experiments, subjects first studied weakly associated cue-target pairs (e.g., mother – *child*) and were then either asked to restudy the pairs or to retrieve the target item from the cue item (e.g., mother – *?*). Experiment 1 found that on a subsequent final recognition test, initial testing of target items increased the false alarm rate to semantic mediators (e.g., *father*) that were presented as lures on that test, relative to initial restudy. Experiment 2 showed that on a subsequent cued-recall test, initial testing of target items increased the probability that subjects correctly recalled the target item (*child*) when a semantic mediator was provided as retrieval cue (*father* –?), relative to initial restudy. These findings support two key assumptions of the elaboration hypothesis: The increase in false alarms for semantic mediators after testing compared to restudying (Experiment 1) may reflect that mediators were more probable to have been activated during the prior test, and the increased effectiveness of semantic mediators as retrieval cues after testing compared to restudying (Experiment 2) may indicate that semantic mediators were more probable to have been linked with the target item during the prior test.

CONTACT Johannes Bartl johannes.bartl@ur.de Department of Experimental Psychology, Regensburg University, Regensburg, 93040, Germany

While the posttesting effect task demonstrates that tests of previously studied material can benefit the material's long-term retention, a growing body of evidence suggests that even tests administered before the actual information is encountered can constitute a powerful learning instrument (Kornell et al., 2009; for a review, see Kornell & Vaughn, 2016). In a typical pretesting effect task, participants initially acquire the study material via either study or pretest trials, which are typically randomly intermixed. On study trials, subjects are presented with a cue item together with a target item (e.g., smoke – *fog*), whereas on pretest trials, they first have to guess the target item from the cue item (e.g., smoke - *?*) before the cue-target pair is shown intact (e.g., smoke – *fog*). On a later memory test, in which participants are shown the cue item (e.g., smoke – *?*) and asked to recall the corresponding target item, recall performance is typically enhanced for pretested target items compared to studied target items. Similar to the posttesting effect, the pretesting effect has been found to be quite robust, and has been observed for different types of study material, including weakly associated word pairs, trivia questions, text passages, and videos, and has been found for different final test formats, including cued-recall tests, fill-in-the-blank tests, multiple-choice tests, and recognition tests (for reviews, see Chan et al., 2018; Kornell & Vaughn, 2016).

A variant of the elaboration account has been proposed to explain the pretesting effect (e.g., Grimaldi & Karpicke, 2012). The assumption is that when subjects attempt to guess the target item during initial pretesting (e.g., smoke - ?), the search for the answer should activate memory representations related to the cue item (again, these can be concepts such as *water*, *opaque*, *white*). These memory representations may become associated with the cue-target pair (smoke – *fog*) once feedback is provided, thus creating a more elaborate memory trace for pretested information. When on a subsequent final test, participants are presented with the cue item and asked to produce the target item, the extra information associated with a pretested pair may become reactivated and promote recall of the target item by serving as a semantic mediator, thereby inducing the pretesting effect. There is some evidence supporting the elaboration account of the pretesting effect. For instance, it has sometimes been found that the pretesting effect largely disappears when unrelated word pairs are applied as study material (e.g., smoke – *animal*; Grimaldi & Karpicke, 2012; Knight et al., 2012). Such a finding is plausible on the basis of the elaboration account since there is no semantic relationship between the semantic mediators and the target item that could promote recall of the target item.

To date, however, more direct evidence is lacking that elaboration processes can contribute to the pretesting effect. A priori, one might assume that Carpenter's (2011) findings may easily generalise from the posttesting-effect task to the pretesting-effect task because both tasks share some similarities, thus providing more definite support for the elaboration account of the pretesting effect. However, this conclusion may be premature. One of the major differences between the posttesting-effect and the pretesting-effect tasks is that the number of overt errors that occur during pretesting is typically much higher than during posttesting, typically close to 100% during pretesting and often in the range between 0% and 50% during posttesting (Rowland, 2014). This commonly observed difference in error rates may well influence how the beneficial memory effects are induced in the two tasks.

For instance, the error-correction account of the pretesting effect provides quite a different view than the elaboration account about how the generation of erroneous guesses during pretesting induces the pretesting effect (Carrier & Pashler, 1992). According to this account, the production of an erroneous guess combined with subsequent corrective feedback leads to an error signal which suppresses the link between the cue item and the incorrect guess and thus increases the relative strength of the link between the cue item and the (correct) target item (Knight et al., 2012). Consequently, the probability that the correct retrieval route is accessed on the final test should be increased, thus creating the pretesting effect. In contrast to the elaboration account, the error-correction account thus leads to the prediction that erroneous guesses are suppressed to facilitate access to the cue-target association, which aligns with connectionist models of error-correction learning (e.g., McClelland et al., 1986). Thus far, only few studies have examined whether error correction plays a critical role also for the pretesting effect. In one of those studies, Seabrooke et al. (2022) found that the size of the pretesting effect was more pronounced for smaller (within-category) errors than greater (between-category) errors, which was argued to be inconsistent with the error-correction account's prediction that larger errors should create a stronger error signal and thus a more pronounced pretesting effect. There is thus little evidence to date to support a role of error correction for the pretesting effect.

The present study

The primary goal of the current study was to put the elaboration account of the pretesting effect to a test. The account assumes that the pretesting effect is induced primarily because pretesting activates information that is semantically linked to the cue item and which becomes linked to the cue-target information, thus promoting later recall of the target information. Three experiments were conducted which were based on the previous study by Carpenter (2011) but used a pretesting procedure instead of a posttesting procedure. In the initial acquisition phase of all three experiments, participants first studied weakly associated cue-target pairs (e.g., mother – *child*), half of which were pretested. Like in Experiment 1 of the

Carpenter study, subjects in our Experiment 1 completed a final item recognition test after the initial acquisition phase which included the target item of each pair as well as new items which, for each target item, included also a semantic mediator (e.g., *father*). In Experiments 2 and 3, final-test performance was measured through cued recall. Similar to Carpenter's Experiment 2, subjects were presented with a retrieval cue from which they needed to generate the target item. For half of the subjects of each experiment, the retrieval cue was the original cue item (mother – *?*) and for the other half of the subjects was a semantic mediator (father – *?*). Experiments 2 and 3 only differed in the length of the retention interval preceding the final test (20 min vs. 4 h).

If the Carpenter (2011) findings generalise from the post-testing to the pretesting procedure, then in Experiment 1 a higher hit rate should arise for target items following pretesting than study only, and in Experiments 2 and 3 a higher correct recall rate for target items should arise following pretesting than study only when the original cue item is used as a retrieval cue, thus reflecting a pretesting effect. More important, a generalisation of the Carpenter findings would also entail an increased false-alarm rate for semantic mediators following pretest than study-only trials (Experiment 1) and an increased recall of target items following pretest than study-only trials when a semantic mediator is provided as a retrieval cue (Experiments 2 and 3).

Such a pattern of results would support the elaboration account of the pretesting effect but would be difficult to explain based on alternative accounts of the pretesting effect like the error-correction account. According to the elaboration account, the activation of extra information during pretesting should make such information more familiar, thus boosting the false alarm rate for potential extra information (i.e., the mediator) on a later test (Experiment 1), and should make it easier to access the target item from the mediator on a later cued-recall test, thus promoting correct recall of the target items (Experiments 2 and 3). On the basis of the error-correction account, however, false alarm rates to mediators or cued recall of target items from mediators should either not differ between the pretesting and study-only conditions, or may even be decreased in the pretesting condition, relative to the study-only condition. The latter may be the case if error correction results not only in suppression of the link between the cue and the guess, but also between the cue and other related information, such as the mediator (see also General Discussion, for a more thorough discussion of alternative accounts of the pretesting effect).

Experiment 1

Method

Participants

To determine sample size in Experiments 1 and 2, a power analysis was conducted using G*Power (version 3.1.9.2; Faul et al., 2009). Based on the metaanalytic effect size estimate of the pretesting effect (Hedge's $g = .44$; Boustani & Shanks, 2022), 43 subjects are required overall when alpha was set to .05 and beta was set to .20. Actual sample size was slightly higher than recommended, with 54 students taking part in Experiment 1 (mean age = 24.1 years; 45 female, 9 male, 0 diverse). All participants spoke German as their native language. All subjects gave their spoken informed consent and received either course credit or a compensatory amount of money for their participation. All reported studies were carried out in accordance with the provisions of the World Medical Association's Declaration of Helsinki.

Materials

Twenty-four weakly associated word pairs (e.g., acrobat – *rope*) drawn from the Nelson et al. (1998) norms were used as study material. The forward association strength was .03 on average (with a range of .01 to .07), which means that subjects would be expected to produce a target item approximately 3% of the time as their (first) associate when provided with a given cue item. Those 24 items were translated into German[1] and divided into two subsets, A and B, consisting of 12 word pairs each. For the one half of the subjects, subset A was assigned to pretest trials while subset B was assigned to study trials (all even-numbered subjects); for the other half of the participants, the assignment of subsets to trial types was reversed (all odd-numbered subjects).

Following Carpenter (2011), the final recognition test consisted of 96 items, 48 of which were old items (the 24 cue items and 24 target items) and 48 of which were new items. Twenty-four of those new items were semantic mediators. In particular, for each of the 24 word pairs, two semantic mediators (i.e., two words with high preexisting associations with the cue item) were drawn also from the Nelson et al. (1998) norms, resulting in 48 mediators overall. These 48 words were divided into two lists, a standard list and a spare list, with each list containing one mediator for each of the 24 cue-target pairs. For example, for the cue item acrobat, the items *gymnast* and *circus* were used as semantic mediators, with forward association strengths of .20 and .17 towards the cue item, but no preexisting association towards the target item *rope*.[2] Mean forward association strength between the cue item and the corresponding mediator items was comparable for the default and spare lists (i.e., .24 for both lists). The mediator that was presented as lure on the final recognition test came from the standard list by default unless the mediator of this list was guessed during initial pretesting. In the case that the mediator was guessed during pretesting, the mediator of the spare list was used on the final test.[3] The 24 remaining new items of the final test were completely unrelated to any of the cue items, target items, or mediator items, and were also drawn from the Nelson et al. (1998) norms. Mean word length did not differ significantly between cue items, target items, unrelated items, and mediator items of the standard and spare lists (mean number of letters: 7.3, 5.8, 6.7, 6.1, 7.1), $F(4, 115) = 1.95$, $MSE = 5.03$, $p = .11$, $\eta_p^2 = .06$. The frequency of occurrences per million words according to the Hyperspace Analogue to Language frequency scale (Balota et al., 2007) did also not differ significantly between the five item types (occurrences per million: 236.2, 326.7, 85.6, 406.5, 235.3), $F(4, 115) = 2.24$, $MSE = 154213.1$, $p = .07$, $\eta_p^2 = .07$. All 96 items of the final test were translated into German.

Design and procedure

The experiment was conducted online via individual meetings, using the videotelephony software program Zoom (Zoom Video Communications), and consisted of three different phases: an acquisition phase, a distractor phase, and a final test phase. During the acquisition phase, participants were presented all 24 word pairs, with half of the cue items presented together with their target item (e.g., acrobat – *rope*) for 10 s each (study-only condition). For the other half of the items, at first only the cue was presented (e.g., divorce – *?*) for 5 s and participants were asked to guess the target item orally, before the complete word pair (e.g., divorce – *lawyer*) was revealed for another 5 s (pretest condition). In both conditions participants were then asked to rate the relatedness between the cue item and the target item on a scale from 1 (*not related at all)* to 5 (*highly related*). Participants gave their ratings orally and were given no time limit. To reduce the probability of ceiling effects on the final test, there was no instruction that the word pairs were to be remembered and would be tested later. The presence of the rating task thus served as an implicit explanation to the subjects as to why the word pairs were presented to them in the first place. All word pairs were presented in a randomised order for each participant.

After the acquisition phase, participants were asked (i) to play an online spot-the-difference task (https://www.suchbilder.com/suchbild-fehlerbild) for 5 min, (ii) to engage in Raven's standard progressive matrices task (Raven et al., 2000) for 12 min, (iii) to play an online version of the video game Tetris (https://www.geo.de/geolino/spiele/13349-rtkl-onlinespiel-tetris) for 3 min. The distractor phase was followed by the final recognition test. In this test, a total of 96 items, consisting of 48 old items (all 24 cue and target items of the acquisition phase) and 48 new items (24 semantic mediators and 24 unrelated items), were presented in random order. Each of the items was read out loud by the experimenter and participants were asked to rate whether the presented item was an old or new item. In particular, a scale ranging from 1 (*definitely new*) to 6 (*definitely old*) was displayed on the screen throughout the test and participants gave their response orally. There was no time limit or corrective feedback during the final test. Experiment 1 thus consisted of a 2×4 design with the within-subjects factors type of practice (study only vs. pretest) and type of test item (target item vs. cue item vs. mediator item vs. unrelated item).

Results

Initial pretest

During the pretest, participants came up with a guess within the 5-s time frame in 96.5% of the trials. They correctly guessed the target items of 4.6% of the item pairs. Those pairs were removed from further analyses since the effect of erroneous guesses on subsequent memory was of main interest.

Final test

Hit rate. Following Carpenter (2011), performance in the final recognition test was analyzed separately for old and new items, each as a function of type of practice (study only vs. pretest). All responses of 4 or higher were considered as "yes" answers and were therefore counted as hits for the old items, since participants correctly classified an old item as old, and as false alarms for new items, since a new item was erroneously classified as old. All analyses were calculated (i) for all "yes" answers (ratings of 4, 5, and 6), (ii) only for "yes" answers given with a medium confidence (ratings of 5 and 6), and (iii) only for "yes" answers given with high confidence (only ratings of 6). Table 1 depicts hit rates for the different confidence levels as a function of type of practice (study only vs. pretest) and type of test item (cue item vs. target item). Pairwise comparisons revealed a pretesting effect – as reflected in a significantly higher hit rate for target items following pretest than study-only trials – both for all "yes" answers (0.90 vs. 0.87), $t(53) = 2.26$, $p = .028$, Cohen's $d = 0.31$, as well as for "yes" answers given with medium confidence (0.87 vs. 0.83), $t(53) = 2.39$, $p = .020$, Cohen's $d = 0.33$, but not for "yes" answers given with high confidence (0.79 vs. 0.77), $t(53) = 1.50$, $p = .14$, Cohen's $d = 0.20$. For cue items, there was a pretesting effect regardless of confidence level, all $ts \geq 3.90$, all $ps < .001$.[4]

False alarm rate. Table 2 depicts false alarms for the different confidence levels as a function of TYPE OF PRACTICE (study only vs. pretest) and TYPE OF TEST ITEM

Table 1. Hit rate (in %) for old items (cues and targets) in Experiment 1 as a function of type of practice and confidence of response. Standard errors are shown in parentheses.

Confidence of response	Study-only	Pretest
All "yes" responses (4s, 5s, 6s)		
Cue items	89.51 (1.59)	96.11 (0.94)
Target items	86.57 (1.58)	90.01 (1.51)
Medium confidence "yes" responses (5s, 6s)		
Cue items	88.12 (1.80)	94.79 (1.17)
Target items	82.72 (1.98)	86.51 (1.84)
High confidence "yes" responses (only 6s)		
Cue items	81.94 (2.44)	90.52 (1.75)
Target items	76.70 (2.32)	79.40 (2.74)

Table 2. False alarms (in %) for new items (mediator items and unrelated items) in Experiment 1 as a function of type of practice and confidence of response. Standard errors are shown in parentheses.

Confidence of response	Study-only	Pretest
All "yes" responses (4s, 5s, 6s)		
Mediator items	4.17 (0.85)	8.39 (1.33)
Unrelated items	2.47 (0.68)	1.70 (0.64)
Medium confidence "yes" responses (5s, 6s)		
Mediator items	2.62 (0.62)	4.79 (0.87)
Unrelated items	1.70 (0.56)	0.93 (0.42)
High confidence "yes" responses (only 6s)		
Mediator items	0.77 (0.40)	1.83 (0.50)
Unrelated items	1.23 (0.46)	0.31 (0.22)

(mediator item vs. unrelated item). A 2 × 2 analysis of variance (ANOVA) of the two within-subjects factors on the false alarm rate for all "yes" answers revealed a significant main effect of TYPE OF TEST ITEM, $F(1,53) = 19.74$, $MSE = 25.48$, $p < .001$, $\eta_p^2 = .27$, with an overall higher false alarm rate for mediator items than for unrelated items (6.3% vs. 2.1%), but no significant main effect of TYPE OF PRACTICE, $F(1,53) = 3.13$, $MSE = 25.48$, $p = .083$, $\eta_p^2 = .06$. More important, there was a significant interaction between the two factors, $F(1,53) = 13.20$, $MSE = 25.48$, $p < .001$, $\eta_p^2 = .20$. Pairwise comparisons revealed that there was a higher false alarm rate in the pretest condition compared with the study-only condition for mediator items (8.4% vs. 4.2%), $t(53) = 3.06$, $p = .004$, Cohen's $d = 0.42$, but not for unrelated items (1.7% vs. 2.5%), $t(53) = 0.80$, $p = .43$, Cohen's $d = 0.11$. The same pattern of results emerged for "yes" answers with medium confidence and "yes" answers with high confidence, again showing significant interactions of the two factors, all $Fs \geq 6.54$, all $ps \leq .013$.

Additional analysis

Like in Carpenter's (2011) earlier study, recognition performance in the present Experiment 1 was analyzed separately for hits and false alarms. We chose this method of analysis since it allowed us to compare the present results as directly as possible with the findings of Carpenter. While the results showed both a higher hit rate for old items and a higher false alarm rate for mediator items in the pretest condition compared to the study-only condition, the question arises whether this pattern of results could be attributed to the use of a more liberal response bias in the pretest than study-only condition, resulting in a higher probability that items presented on the final recognition test were classified as old – producing both more hits for target (and cue) items and more false alarms for mediator items. To examine this possibility, we conducted an ROC (receiver operating characteristics) analysis to estimate the response biases for both study-only and pretest trials. To this end, the proportion of correctly recognised target items (i.e., the hit rate) and the proportion of incorrectly recognised lure items (i.e., the false alarm rate) were cumulated across the rating scale starting at the most confident criterion, i.e., definitely old ("1"). This method leads to an empirical ROC curve that relates the hit and false alarm rates for different response criteria (e.g., Macmillan & Creelman, 2004). Using this 6-point scale, hit and false alarm rates were obtained for five different response criteria.

In the next step, the recognition data were analyzed using a signal detection approach. We assumed unequal variance for the distribution of old and new items to account for the typically asymmetric shape of the ROC and the data were described using the unequal variance signal detection model (e.g., Wixted, 2007). This model bases recognition judgments on a single source of information, i.e., the overall memory strength of the items. If an item exceeds the response criterion c_i, which is associated with a certain confidence level i, but not the criterion c_{i-1}, participants rate the item as i accordingly. The memory strength of old items compared to new items – i.e., sensitivity – can be derived from the distance between the mean values of the underlying strength distributions of the old and new items (d'). Applying the model to our 5-point ROC data results in seven free parameters (memory strength of the old items d_a, variance of the distribution of the old items λ, and five criterion points c_1–c_5) and consequently three degrees of freedom when testing the goodness of fit of the model. To estimate the model parameters, we used maximum likelihood methods, which can also be used for statistical tests.

Table 3 shows the statistics of goodness of fit and maximum likelihood estimates of the model's parameters d′, λ, and c_1 to c_5 for study-only and pretest trials. The unequal-variance signal detection model provided a good fit to the recognition data of the two types of practice (study only vs. pretest), all $\chi^2s(3) \leq 1.83$, all $ps \geq 0.59$. Neither the numerical difference in d' (i.e., sensitivity) nor in λ between study-only and pretest trials reached significance, all $\chi^2s(1) \leq 0.56$, all $ps \geq 0.45$. The finding that sensitivity was similar for the study-only and pretest conditions converges with the above observation that pretesting increased *both* hit rates for target items and false alarm rates for mediator items. Statistical analysis of the model parameters further showed that the estimated response criteria c_1–c_4 were significantly reduced – i.e., more liberal – in the pretest condition than the study-only condition, all $\chi^2s(1) \geq 6.76$, all $ps \leq 0.009$. This analysis suggests that in the pretest condition, subjects showed a stronger tendency to categorise items as old, regardless of whether the item had actually been encountered during acquisition.

Discussion

The results of Experiment 1 show a pretesting effect for old items, as reflected in higher hit rates for pretest than for study-only target items. More important, Experiment 1 revealed a higher false alarm rate for (new) mediator items, but not for (new) unrelated items in the pretest

Table 3. Fit of the unequal-variance signal detection model to the recognition data of Experiment 1.

	Parameter estimates							Goodness of fit		
Study format	d''	λ	c_1	c_2	c_3	c_4	c_5	χ^2	df	p
Study only	3.88	2.01	2.42	1.96	1.71	1.24	0.45	1.93	3	0.59
Pretest	3.64	1.82	2.12*	1.66*	1.36*	1.06*	0.45	1.22	3	0.75

Note: d' = general memory strength; λ = variance of the target distribution; c_{1-5} = response criteria; df = degrees of freedom; * Significant deviations from performance in the study-only condition ($p < .05$).

condition, compared to the study-only condition. These results generalise the findings of Carpenter (2011, Experiment 1) from the posttesting to the pretesting procedure and support the elaboration account of the pretesting effect by providing a first demonstration that pretesting may increase the probability that semantic mediators are activated. [5] Specifically, activation of the semantic mediators during the pretest may not only increase the strength of memory representations of the target items – by improving the integration of the target items with the cue item – but also the strength of memory representations of the mediator items.[6] Strengthening both the target and mediator items may reinforce subjects' impression on the final recognition test that these items had been encountered before, which could lead them to adopt a more liberal response bias in the pretest condition than the study-only condition (see also, Spitzer & Bäuml, 2009, who reported similar effects of testing on response bias in the retrieval-practice task).

The goal of Experiment 2 was to examine whether semantic mediators are not only more probable to be activated during the pretest but are also more probable to be associated with the target items as a result of pretesting compared to study only. To test this prediction, Experiment 2 examined whether presenting a semantic mediator as a retrieval cue on the final test produces the target item more effectively after initial pretesting than after study only.

Experiment 2

Method

Participants

Closely following the recommendation of the power analysis reported in Experiment 1, 40 students were recruited for each of the two between-subjects conditions of Experiment 2, resulting in 80 students overall (mean age = 22.8 years; 67 female, 12 male, 1 diverse). All participants spoke German as their native language and reported no neurological or psychiatric disease. All subjects gave spoken informed consent and received either course credit or a compensatory amount of money for their participation.

Materials, design, and procedure

Experiment 2 was identical to Experiment 1 with the only exception that the final test consisted of a cued-recall task. At the beginning of this test, all subjects were informed that the experimenter would read aloud words which either could be the cue items from the acquisition phase or new items. Subjects were instructed that whenever one of the words read aloud reminded them of a target item from the acquisition phase (e.g., *rope*), they should say this target item aloud, regardless of whether the word read aloud was a familiar or was a new item. For half of the subjects, only the original cue items were presented as retrieval cues on the final test (e.g., acrobat), whereas for the remaining half of the subjects, only the semantic mediators (e.g., gymnast) were presented as retrieval cues on the final test. Experiment 2 thus consisted of a 2 × 2 design with the within-subjects factor type of practice (study only vs. pretest) and the between-subjects factor type of cue (original vs. mediator). There was no time limit or corrective feedback during the final test.

Results

Initial pretest

During the pretest, participants came up with a guess within the 5-s time frame in 98.7% of the trials. They correctly guessed the target items of 4.6% of the item pairs. Like in Experiment 1, those pairs were removed from further analyses.

Final test

Correct recall

Table 4 shows mean recall rates on the final cued-recall test as a function of TYPE OF PRACTICE (study only vs. pretest) and TYPE OF CUE (original vs. mediator). A 2 × 2 ANOVA with the two factors revealed main effects of TYPE OF PRACTICE, $F(1,78) = 19.66$, $MSE = 111.15$, $p < .001$, $\eta_p^2 = .20$, and TYPE OF CUE, $F(1,78) = 592.95$, $MSE = 177.50$, $p < .001$, $\eta_p^2 = .88$, reflecting overall greater recall of target items from pretest than study-only trials (68.8% vs. 61.5%) and an overall decrease in recall if, on the final test, mediator items were used as retrieval cues instead of the original cue items (39.5% vs. 90.8%). These main effects were qualified by a statistically significant

Table 4. Correct recall (in %) in Experiments 2 and 3 as a function of type of practice and type of cue. Standard errors are shown in parentheses.

	Study-only	Pretest
Experiment 2		
Original cue	90.63 (1.27)	90.98 (1.65)
Mediator cue	32.29 (2.07)	46.72 (2.41)
Experiment 3		
Original cue	77.08 (3.22)	81.10 (2.92)
Mediator cue	27.92 (2.28)	45.02 (2.35)

interaction between the two factors, $F(1,78) = 17.83$, $MSE = 111.15$, $p < .001$, $\eta_p^2 = .19$. Consistently, pairwise comparison showed that recall of target items did not differ between study-only and pretest trials when the original cue was presented as retrieval cue (91.0% vs. 90.6%), $t(39) = 0.19$, $p = .85$, Cohen's $d = 0.03$, revealing no pretesting effect, whereas recall of target items was greater for pretest than study-only trials when the mediator was provided as retrieval cue (46.7% vs. 32.3%), $t(39) = 5.26$, $p < .001$, Cohen's $d = 0.83$, demonstrating a pretesting effect of 14.4% for this type of retrieval cue.

Intrusions. All incorrect answers participants produced during the final test were counted as intrusions. A 2 × 2 ANOVA with the factors TYPE OF PRACTICE (study only vs. pretest) and TYPE OF CUE (original vs. mediator) revealed a main effect of TYPE OF CUE, $F(1,78) = 41.36$, $MSE = 2.36$, $p < .001$, $\eta_p^2 = .35$, reflecting that, overall, fewer intrusions were made for the original cues than for the mediator cues (0.4 vs. 2.0 intrusions). Neither the main effect of TYPE OF PRACTICE, $F(1,78) = 3.82$, $MSE = 1.19$, $p = .054$, $\eta_p^2 = .05$, nor the interaction between the two factors, $F(1,78) = 1.51$, $MSE = 1.19$, $p = .22$, $\eta_p^2 = .02$, reached statistical significance.

Discussion

The results of Experiment 2 showed that a reliable pretesting effect can arise when a semantic mediator is presented as retrieval cue on the final test, as reflected in greater recall of target items in the pretest than study-only condition. Like Experiment 1, these results generalise the Carpenter (2011, Experiment 2) findings from the posttesting to the pretesting procedure. Critically, however, no standard pretesting effect arose in Experiment 2 when the original cue item was presented as retrieval cue on the final test. Since recall rates were already very high in the study-only (control) condition (i.e., ca. 91%), there was not much room left for improvement for the pretesting condition. Therefore, the goal of Experiment 3 was to apply the same task as in Experiment 2, but to make the final test more difficult – by increasing the length of the retention interval preceding the final test to 4 h – to reduce recall rates in the study condition and provide more room for a potential benefit of pretesting on later retention.

Experiment 3

Method

Participants

Closely following the power analysis reported in Experiment 1, 80 students (mean age = 24.1 years; 47 female, 33 male, 0 diverse) took part in Experiment 3. All participants spoke German as their native language and gave their spoken informed consent. In return for their participation, all subjects received either course credit or a monetary award.

Materials, design, and procedure

Materials as well as procedural details were identical to Experiment 2 except for the retention interval. While in Experiment 2, the final cued-recall test took part after a 20 min distractor phase, subjects in Experiment 3 were finished for the time being after the study phase, but were asked to complete the final test on a second appointment 4 h later. All subjects who attended the first session did also complete the second session.

Results

Initial pretest

During the pretest, participants were able to come up with a guess within the 5-s time frame in 95.1% of the trials. They correctly guessed 4.6% of the response words. Those pairs were removed from further analyses.

Final test

Correct recall. Mean recall rates on the final cued-recall test as a function of TYPE OF PRACTICE (study only vs. pretest) and TYPE OF CUE (original vs. mediator) are shown in Table 4. A 2 × 2 ANOVA of the two factors revealed main effects of TYPE OF PRACTICE, $F(1,78) = 84.94$, $MSE = 52.49$, $p < .001$, $\eta_p^2 = .52$, and TYPE OF CUE, $F(1,78) = 134.32$, $MSE = 541.01$, $p < .05$, $\eta_p^2 = .63$, reflecting overall greater recall of target items from pretest than study-only trials (63.1% vs. 52.5%) and an overall decrease in recall if, on the final test, mediator items were used as retrieval cues instead of the original cue items (36.5% vs. 79.1%). These main effects were qualified by a statistically significant interaction between the two factors, $F(1,78) = 32.64$, $MSE = 52.49$, $p < .001$, $\eta_p^2 = .30$. Pairwise comparison showed that recall of target items was greater for pretest than study-only trials both when the original cue was presented as retrieval cue (81.1% vs. 77.1%), $t(39) = 3.32$, $p = .002$, Cohen's $d = 0.53$, resulting in a pretesting effect of 4.0%, and when the mediator was presented as retrieval cue (45.0% vs. 27.9%), $t(39) = 8.78$, $p < .001$, Cohen's $d = 1.39$, revealing a pretesting effect of 17.1%.

Intrusions

A 2 × 2 ANOVA with the factors TYPE OF PRACTICE (study only vs. pretest) and TYPE OF CUE (original vs. mediator) revealed main effects of TYPE OF PRACTICE, $F(1,78) = 15.98$, $MSE = 1.06$, $p < .001$, $\eta_p^2 = .17$, and of TYPE OF CUE, $F(1,78) = 23.00$, $MSE = 4.32$, $p < .001$, $\eta_p^2 = .23$, reflecting that, overall, pretest trials led to fewer intrusions than study-only trials (1.6 vs. 2.2 intrusions) and intrusions were lower for the original cues than for the mediator cues (2.7 vs .1.1 intrusions). There was no significant interaction between the two factors, $F(1,78) = 1.51$, $MSE = 1.06$, $p = .22$, $\eta_p^2 = .02$.

Discussion

The results of Experiment 3 again showed that using a semantic mediator as a retrieval cue can lead to a reliable pretesting effect. Unlike Experiment 2, the results of Experiment 3 also demonstrated a typical pretesting effect when the original cue item was used as retrieval cue. In fact, increasing the length of the retention interval from 20 min (Experiment 2) to 4 h (Experiment 3) decreased recall performance in the study-only condition by roughly 14 percentage points and thus provided more room for a recall enhancement in the pretest condition, relative to the study-only condition. The findings of Experiment 3 again align with the elaboration account of the pretesting effect, supporting the assumption that pretesting increases the probability that extra information such as the semantic mediator is linked with the target item and thus facilitates retrieval of the target item on a later test.

The argument could be made that in Experiments 2 and 3, pretests again may have caused a shift towards a more liberal response bias, thus creating the observed benefit of pretesting when the mediator item was presented as retrieval cue on the final test. If so, we would expect to observe more responses given overall and thus a higher rate of correct recall as well as a higher number of intrusions (incorrect responses) in the pretest condition than the study-only condition. In contrast, in the present Experiments 2 and 3, a higher rate of correct recall and a *lower* number of intrusions (i.e., numerically fewer intrusion in Experiment 2 and statistically fewer intrusions in Experiment 3) were produced in the pretest condition, which suggests that final test format might play a critical role for whether pretesting leads to a more liberal response bias. Future work is required to examine the specific conditions under which pretesting modulates response bias.

General discussion

The results of the present Experiments 1 and 3 demonstrate that pretesting of target items can lead to better memory performance of those items on a later retention test, regardless of whether a recognition or cued-recall format is applied, thus replicating prior work (e.g., Kornell et al., 2009; Seabrooke et al., 2021). Experiment 2 failed to find a pretesting effect on a final cued-recall test, which was probably due to a ceiling effect. More important, the results of the present three experiments extend Carpenter's (2011) findings from the posttesting effect to the pretesting effect task, by suggesting that an analogous pattern of results can arise for both tasks when semantic mediators are part of a final retention test: In Experiment 1, subjects who were pretested on target items (e.g., mother – ?) were later more probable to make false alarms to a semantic mediator (e.g., father) – i.e., an item that was never shown during acquisition but closely related with the cue item – than were participants who were given additional time to study the cue-target pair (mother – *child*). Experiments 2 and 3 provide a first demonstration that the pretesting effect can also occur when a semantic mediator (e.g., father) is used as a retrieval cue on the final test, instead of the original cue item of a pair (mother). The current experiments thus add to the parallels between the posttesting and pretesting procedures. Indeed, both posttesting and pretesting have been found yet to be able to reduce time-dependent-forgetting (e.g., Kliegl et al., in press a; Roediger & Karpicke, 2006), forgetting caused by retroactive interference (Halamish & Bjork, 2011; Kliegl et al., 2023; Potts & Shanks, 2012), and to be particularly beneficial when multiple initial test cycles are applied (Karpicke & Roediger, 2007; Kliegl et al., in press b).

The present findings align with the elaboration account of the pretesting effect. The account assumes that pretesting leads to the activation of memory representations related to the cue item, which can be used as semantic mediators on a final test through which the target item is retrieved. If pretesting creates a more elaborate memory trace than does additional study since it leads to the activation of information related to the cue item, then it should become more probable that this information will appear familiar to participants on a final recognition test. Consequently, the false alarm rate for mediators should be higher in the pretest condition than in the study-only condition, as was found in Experiment 1. Furthermore, based on the account's assumption that pretesting promotes the linking of the semantic mediator with the target item, the presentation of a semantic mediator as a retrieval cue on a final cued-recall test should make it more probable that the target item is correctly recalled, which is exactly what the results of Experiments 2 and 3 show.

In contrast, the findings do not align with the error-correction account (Carrier & Pashler, 1992). According to this account, the production of errors during pretesting and the subsequent correction of these errors by immediate feedback should weaken the incorrect retrieval paths between the cue and the initial guess and, as a result, the relative strength of the correct retrieval path between the cue and the target should be increased. Based on this account, false alarm rates to mediators and recall of target items from cue items should either not be affected by the initial practice format (pretest vs. study only) or may even be reduced in the pretest condition, compared to the study-only condition, if the suppression mechanism also weakens the path between the cue item and extra information, such as the mediator. However, the present findings do not align with either of the two possibilities, which also fits with other prior work suggesting that the role of error correction for the pretesting effect may be negligible (e.g., Grimaldi & Karpicke, 2012; Seabrooke et al., 2022).

While the elaboration and the error-correction accounts are among two of the most prominent explanations of the pretesting effect, several further accounts are often discussed in the literature, such as the episodic-context account, the attentional account, and the search-set account. The episodic-context account (e.g., Metcalfe & Huelser, 2020; Overman et al., 2021) is based on the assumption that during the encoding of information, features of the temporal context that are present while encountering the to-be-studied information are also stored. Pretested items could therefore be associated with an enriched and more distinct temporal context due to the integrated memory representation of the pretest context together with the study context, whereas the study items should only be associated with the study context. Since the episodic-context account thus assumes that the pretesting effect is driven by the presence of additional contextual features, semantic mediators should not play a critical role for the effect. Consequently, the present findings are difficult to explain on the basis of this account alone.

The attentional account explains the pretesting effect by assuming that pretesting enhances attentional encoding of the target item when it is revealed at the end of a pretest trial, thus promoting recall of the target item on a subsequent final test (e.g., Potts & Shanks, 2014; Zawadzka & Hanczakowski, 2019). This account assumes that pretesting boosts the encoding of the target item itself, without strengthening the link between the cue and the target items (e.g., Potts et al., 2019; Seabrooke et al., 2021). If the enhanced encoding exclusively affects the target item and no other items, then subject's responses to semantic mediators on a final recognition test or on a cued-recall test should not differ as a function of whether the target items were initially pretested or studied only. Therefore, the predictions of the attentional account also appear difficult to reconcile with the present findings.

Finally, the search-set account (e.g., Grimaldi & Karpicke, 2012) assumes that when subjects attempt to guess a correct response when provided with a cue item (e.g., smoke) on a pretest trial, many related items become activated, such as *fire*, *steam*, *fog*, or *water*. While the subject will probably make an incorrect overt guess during pretesting (such as *water*), the correct target (fog) should also have been activated as a member of the search set of related concepts and should thus lead to enhanced encoding of that target item when the correct answer feedback is subsequently provided. Since, similar to the attentional account, the account assumes that the enhanced encoding boosts only later memory of the target item, subjects' responding to semantic mediators on the final test should not be affected by initial study format. The present findings thus do not easily align with the search-set account.

Although the present study was primarily concerned with investigating the cognitive processes underlying the pretesting effect, some preliminary conclusions for the use of pretests as a teaching tool in educational contexts can also be drawn from the results. In particular, the present findings suggest that pretests can produce both negative and positive transfer effects. The results of Experiment 1 suggest that pretesting may induce a negative transfer effect when it is used as a learning tool, since it may sometimes increase a student's propensity to falsely recognise previously unstudied information as "old" when the unstudied information is closely related to the pretested information. The results of Experiment 2 suggest a positive transfer effect, as pretesting could improve a student's ability to retrieve previously studied information even when the cues (e.g., test questions) differ between the pretest and the final test. Certainly, further research on the generalizability and boundary conditions of these two transfer effects would be crucial when it comes to assessing the role of pretests as a learning strategy in education.

To *conclude*, the results of the three present experiments showed that presenting semantic mediators as lures on a final recognition test can lead to more false alarms for the mediators following initial pretesting than study only (Experiment 1) and presenting semantic mediators as retrieval cues on a final cued-recall test lead to higher correct recall of target items following initial pretesting than study only (Experiments 2 and 3). These findings emphasise the role of semantic mediators for the pretesting effect and thus provide direct support for the elaboration account of the pretesting effect but are more difficult to reconcile with alternative explanations of the pretesting effect, such as the error-correction account.

Notes

1. Since the study material consisted of German translations of word pairs that were standardized with English materials, the association strengths between cue and target items of these translations may not perfectly match the association strengths of the original English version of the materials.
2. This approach follows the specifications of Carpenter (2011). Indeed, previous research has shown that a (reduced) pretesting effect can occur even when there is no clear relationship between the participant's guess during the pretest (which is often also likely to be a mediator item) and the target item (Cyr & Anderson, 2018).
3. To make sure that, across all subjects, mean forward association strength between the cue and mediator items was comparable for the study and pretest conditions, the following proceeding was applied: If participant X guessed the mediator of the standard list (e.g., *gymnast*) during pretesting, the mediator of the spare list (e.g., *circus*) was used on the final recognition test, as already pointed out. Critically, for the subsequent participant (X + 1), the same mediator item then was used as the mediator item for the study condition. Across subjects, all mediator items thus were presented equally often as pretest and study mediator items in the final test.
4. The cue items served as filler items on the final recognition test to obtain as many "old" items as "new" items (i.e., 48 items each). Since we were primarily interested in the hit rates of the target items, we decided to report the hit rates of the

cue items separately, and not as part of an ANOVA (see also Carpenter, 2011).

5. However, one difference between the present findings and the Carpenter findings is that in the earlier study, the magnitude of the posttesting effect was most pronounced at the highest confidence level, while in the present study, level of confidence did not appear to have an effect on the size of the effect. The reason for these minor discrepancies in results may not necessarily be due to the fact that the earlier study employed a posttesting procedure and the present study used a pretesting procedure but could also be the result of other procedural differences between studies (i.e., Carpenter had subjects study fewer word pairs overall during the initial acquisition phase and informed subjects about the presence of a final test).
6. While pretest-induced elaboration processes may increase the memory strengths of both target and mediator items, ROC analysis is not able to unambiguously determine whether such processes occur with the present experimental setup.

Open Scholarship

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The data and materials are openly accessible at https://osf.io/5vdu4/ and https://osf.io/5vdu4/ .

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) awarded to Oliver Kliegl [grant number: KL 2967/2-1].

Open practices statement

The data and materials for all experiments are available at https://osf.io/5vdu4/, and none of the experiments was preregistered.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Boustani, S., & Shanks, D. R. (2022). Heterogeneity and publication bias in research on test-potentiated new learning. *Collabra: Psychology*, *8*(1), 319–396. https://doi.org/10.1525/collabra.31996

Brady, T. F., Robinson, M. M., Williams, J. R., & Wixted, J. T. (2023). Measuring memory is harder than you think: How to avoid problematic measurement practices in memory research. *Psychonomic Bulletin & Review*, *30*(2), 421–449. https://doi.org/10.3758/s13423-022-02179-w

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 15471552. https://doi.org/10.1037/a0024140

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 633–642. https://doi.org/10.3758/BF03202713

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*(11), 1111–1146. https://doi.org/10.1037/bul0000166

Cyr, A. A., & Anderson, N. D. (2018). Learning from your mistakes: Does it matter if you're out in left foot, I mean field? *Memory*, *26*, 1281–1290. https://doi.org/10.1080/09658211.2018.1464189

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

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

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 801–812. https://doi.org/10.1037/a0023219

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Cognitive psychology of memory, Vol. 2 of learning and memory: A comprehensive reference* (pp. 1–26). Amsterdam, the Netherlands: Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D., Roediger, L.H., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162. https://doi.org/10.1016/j.jml.2006.09.004

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2023). The pretesting effect thrives in the presence of competing information. *Memory*, *31*(5), 705–714. https://doi.org/10.1080/09658211.2023.2190568

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (in press a). The pretesting effect comes to full fruition after prolonged retention interval. *Journal of Applied Research in Memory And Cognition*, https://doi.org/10.1037/mac0000085

Kliegl, O., Bartl, J., & Bäuml, K. H. T. (in press b). Repeated guessing attempts during acquisition can promote subsequent recall performance. *Journal of Experimental Psychology: Applied*, https://psycnet.apa.org/doi/10.1037xap0000493.

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language*, *66*(4), 731–746. https://doi.org/10.1016/j.jml.2011.12.008

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

Kornell, N., & Vaughn, K. E. (2016). How retrieval attempts affect learning: A review and synthesis. *Psychology of Learning and Motivation*, *65*, 183–215. https://doi.org/10.1016/bs.plm.2016.03.003

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Inc.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & T. P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 365–422). MIT Press.

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, *138*, 107296. https://doi.org/10.1016/j.neuropsychologia.2019.107296

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://w3.usf.edu/FreeAssociation/.

Overman, A. A., Stephens, J. D., & Bernhardt, M. F. (2021). Enhanced memory for context associated with corrective feedback:

Evidence for episodic processes in errorful learning. *Memory*, *29*(8), 1017–1042. https://doi.org/10.1080/09658211.2021.1957937

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. (2012). Can testing immunize memories against interference? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1780–1785. https://doi.org/10.1037/a0028218

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

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. Psychology Press. https://doi.org/10.1006/cogp.1999.0735

Roediger, H. L., III, & 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. 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

Seabrooke, T., Mitchell, C. J., Wills, A. J., & Hollins, T. J. (2021). Pretesting boosts recognition, but not cued recall, of targets from unrelated word pairs. *Psychonomic Bulletin & Review*, *28*(1), 268–273. https://doi.org/10.3758/s13423-020-01810-y

Seabrooke, T., Mitchell, C. J., Wills, A. J., Inkster, A. B., & Hollins, T. J. (2022). The benefits of impossible tests: Assessing the role of error-correction in the pretesting effect. *Memory & Cognition*, *50*(2), 296–311. https://doi.org/10.3758/s13421-021-01218-6

Spitzer, B., & Bäuml, K. H. (2009). Retrieval-induced forgetting in a category recognition task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 286–291. https://doi.org/10.1037/a0014363

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152–176. https://doi.org/10.1037/0033-295X.114.1.152

Zawadzka, K., & Hanczakowski, M. (2019). Two routes to memory benefits of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(10), 1748–1760. https://doi.org/10.1037/xlm0000676