Memory

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

The pretesting effect thrives in the presence of competing information

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

To cite this article: Oliver Kliegl, Johannes Bartl & Karl-Heinz T. Bäuml (2023) The pretesting effect thrives in the presence of competing information, Memory, 31:5, 705-714, DOI: 10.1080/09658211.2023.2190568

To link to this article: https://doi.org/10.1080/09658211.2023.2190568

Published online: 16 Mar 2023.

MEMORY
2023, VOL. 31, NO. 5, 705–714
https://doi.org/10.1080/09658211.2023.2190568

The pretesting effect thrives in the presence of competing information

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

Department of Experimental Psychology, Regensburg University, Regensburg, Germany

ABSTRACT

Taking a pretest (e.g., blanket – ?) before some target material (blanket – *sheet*) is studied can promote recall of that material on a subsequent final test compared to material which was initially only studied. Here, we examine whether such pretesting can shield the tested material from interference-induced forgetting, which often occurs when before final testing, related material is encountered. We applied a typical pretesting task but asked subjects, between acquisition and final testing of the target list (list 1), to study two additional lists of items with either completely new and unique pairs (e.g., atom – *cell*) or overlapping – and thus potentially interfering – pairs (e.g., blanket – *sleep*). Target-list recall on the final test showed a typical pretesting effect for unique pairs, but the size of the effect even increased for overlapping pairs, as recall of study-only pairs was impaired, whereas recall of pretest pairs was left largely unaffected. This held regardless of whether a low (Experiment 1) or high (Experiment 2) degree of learning was induced for the interfering material, suggesting that pretesting can indeed protect the tested material from interference. These findings indicate that pretesting could play a significant role in educational settings where information often needs to be retained in the presence of competing information.

ARTICLE HISTORY
Received 7 November 2022
Accepted 7 March 2023

KEYWORDS
Testing effect; pretesting effect; retroactive interference; elaboration

Failure to retrieve previously encoded information from memory does not mean that the information sought has been lost from memory. Rather, research over the last 120 years has shown that such failure of memory often occurs when, prior to the retrieval attempt, additional nontarget information related to the target information has been studied (for reviews, see Kliegl & Bäuml, 2021; Wixted, 2004). One prominent form of this so-called interference-induced forgetting is retroactive interference (RI; Müller & Pilzecker, 1900), which is typically observed when between study and test of the target information, related nontarget information is encountered.

Research has often used the paired-associate learning task to examine RI buildup. In a typical task, study of a target list of stimulus-response word pairs (e.g., blanket – *pillow*) is followed by study of an interpolated list consisting of either overlapping pairs with the same stimulus word but a new response word (e.g., blanket – *sheet*) or unique pairs with completely new stimulus and response words (e.g., tea – *leaves*). When memory of the target list's response words is later assessed (e.g., blanket – *?*), recall of those items is typically found to be impaired for overlapping pairs compared to unique pairs, thus reflecting RI buildup (e.g., Greeno, 1964; Postman & Underwood, 1973).

Prior research shows that several methods can be employed to reduce the detrimental effects of RI on memory performance (for a review, see Kliegl & Bäuml, 2021). One such method involves taking a posttest, that is, retrieval practice, on the initially studied target list before the subsequent nontarget list is studied. For instance, Halamish and Bjork (2011) had participants first study a target list consisting of stimulus-response pairs (e.g., tea - *leaves*) and then asked them to practice the list via either restudy (e.g., tea - *leaves*) or retrieval practice (e.g., tea – *?*) before an interpolated list consisting of either unique or overlapping stimulus-response pairs was studied. While a later test taken on the initial list's response words showed a pronounced decrease in recall from unique to overlapping pairs following initial restudy, thus suggesting typical RI buildup, the decrease in recall from unique to overlapping pairs was much less pronounced after initial retrieval practice, thus suggesting a posttest-induced RI reduction (for similar findings, see Potts & Shanks, 2012).

Posttests can have further benefits for memory, besides the reduction in RI (for reviews, see Bäuml & Kliegl, 2017; Karpicke, 2017; Roediger III & Butler, 2011). For instance, posttests have been shown to reduce forgetting over time (Roediger III & Karpicke, 2006), improve transfer of knowledge (e.g., Butler, 2010) and decrease mind wandering (e.g., Szpunar et al., 2013). However, also pretests administered before the material is studied followed by an opportunity to study the material can enhance retention

CONTACT Oliver Kliegl oliver.kliegl@ur.de Department of Experimental Psychology, Regensburg University, Regensburg 93040, Germany

(for a review, see Kornell & Vaughn, 2016). In a typical task used to examine this pretesting effect, subjects are either asked to study stimulus-response pairs (e.g., blanket – *pillow*) or they are initially shown only the stimulus word for a few seconds and then have to provide a guess as to the associated response word (blanket – *?*) before the stimulus-response pairing is shown intact. On a later final test, in which the response words have to be recalled from the stimulus words (blanket – *?*), recall performance is typically improved for response words that were initially pretested, compared to response words that were initially only studied (e.g., Grimaldi & Karpicke, 2012; Kornell et al., 2009). The pretesting effect has been observed across a variety of study materials and in both lab-based studies and educational settings (for reviews, see Chan et al., 2018; Kornell & Vaughn, 2016). Moreover, the pretesting effect has been shown to parallel the classic testing effect by reducing forgetting over time (Kliegl et al., in press). However, no study has yet investigated whether pretesting also protects memory from RI. This study is the first one in the literature to address the issue.

The pretesting effect is often explained on the basis of the elaboration hypothesis (Carpenter, 2009, 2011; Cyr & Anderson, 2012, 2018; Huelser & Metcalfe, 2012). This account assumes that during an initial pretest which provides a stimulus word and has subjects guess the response word (plate - *?*), memory representations related to the stimulus word become activated, such as *table*, *meal*, *knife*. These memory representations are assumed to become associated with the stimulus-response pair once feedback is provided, thus creating a more elaborate memory trace than for study trials which, due to their more passive nature, may barely involve activation of any associated memory representations. If, on a subsequent final test, the stimulus word is shown and the response word has to be recalled, the memory representations associated with pretested pairs may become re-activated and facilitate retrieval of the response word, thus creating the pretesting effect. Support for the elaboration hypothesis stems from prior work showing that pretests not only improve later recall of the response word but also slow down its retrieval (Huelser & Metcalfe, 2012), since attempting to produce an initially pretested pair may first lead to the retrieval of mediating information through which the response word is ultimately accessed. Such an additional step in the retrieval process would require more time.

On the basis of the elaboration hypothesis – together with the so-called recursive reminding framework (e.g., Jacoby & Wahlheim, 2013; Wahlheim, 2015) – the expectation arises that pretesting should also shield the tested material from the detrimental effects of RI. Indeed, since pretesting is assumed to create a more elaborate memory trace for the pretested stimulus-response pairs, it may become more probable that during the subsequent study of overlapping pairs, subjects are reminded of the initial (target) response word than in the absence of pretesting. In particular, the reminding is assumed to lead to a more integrated memory representation including all responses together with information about the temporal order of occurrence during acquisition. Consequently, subjects' propensity to confuse the target response word with the subsequently studied response words may be reduced at the later test, thus limiting possible interference effects. As a result, the size of the pretesting effect may increase in the presence of competing information, thus paralleling the effects of posttesting on RI. This prediction of the combined elaboration and recursive-reminding accounts has not been examined to date.

The present study

The goal of the present study was to investigate whether pretests can shield the tested material from the detrimental effects of RI. To this end, the present two experiments used a paired-associate learning task in which participants were either shown intact stimulus-response pairs – such as blanket – *pillow* – to study for a later test, or were first only shown the stimulus word and asked to guess the response word – such as blanket – *?* – before the intact stimulus-response pair was revealed and could be studied. Following study of this target list (list 1), two nontarget lists (lists 2 and 3) were presented, which, for half of the pretested and study-only pairs each, consisted of overlapping pairs (e.g., blanket – *sheet*), and, for the remaining half of the pretested and study-only pairs each, consisted of unique pairs (e.g., tea – *leaves*). Following this acquisition phase, participants engaged in a final test on list-1 pairs, in which they were provided with the stimulus words and asked to produce the response words. Both correct recall and intrusions were measured. While lists 2 and 3 were presented only once during acquisition in Experiment 1, the lists were studied twice in immediate succession in Experiment 2. This was done to induce a higher degree of learning for lists 2 and 3 and thus potentially also higher levels of RI on list-1 recall on the final test (e.g., Barnes & Underwood, 1959; Briggs, 1957). Otherwise, all procedural details of Experiments 1 and 2 were identical.

We made three predictions for list-1 recall performance on the final test. First, a typical pretesting effect was expected for unique pairs, as reflected in superior recall performance for pretested than studied responses. Since study of unique list-2 and list-3 pairs should not induce much RI on the final test, list-1 recall should be largely unaffected by the interpolated lists. Second, consistent with the RI literature, it was anticipated that recall of study-only responses would suffer in overlapping pairs compared to unique pairs, thus reflecting typical buildup of RI. Third, and most important, it was expected that such RI would be reduced for pretested pairs. Indeed, pretesting may make it more probable that during study of the overlapping list-2 and list-3 pairs, subjects are reminded of the list-1 response words, enabling the integration of the competing list-2 and list-3 responses, and

diminishing RI effects. The present study thus will shed light on whether the benefit of pretesting on later retention can increase in learning environments where competing information is present.

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 meta-analytic effect size estimate of the pretesting effect (Hedge's $g = 0.44$, Boustani & Shanks, 2022), 43 subjects were required overall when α was set to .05 and β to .20. Closely following this recommendation, 40 students (mean age = 24.6 years) were recruited to participate in Experiment 1. 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. All reported experiments were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Material

Fourty weakly associated stimulus-response pairs (e.g., plate – *fork*; tradition – *christmas*) were drawn from the Nelson et al. (1998) norms and assigned to list 1 – the target list. The forward-association strength of list-1 word pairs ranged from .048 to .067 and was .052 on average, meaning that subjects in the Nelson et al. study produced a given response word about 5% of the time as their first associate when presented with a given stimulus word. For each subject, one half of the 40 pairs were assigned to pretest trials and the remaining half were assigned to study-only trials. Across subjects, all list-1 word pairs were used equally often in pretest and study trials.

For each of the 40 list-1 pairs, another set of 40 pairs were drawn from Nelson et al. for which the stimulus word remained the same but the response word was replaced (e.g., plate – *glasses*). For each subject, 20 of these overlapping pairs were assigned to list 2. In addition, another 40 pairs were drawn from Nelson et al. for which both the stimulus and response words were completely new (e.g., atom – *cell*). For each subject, 20 of these unique pairs were assigned to list 2. Across subjects, for a given list-1 pair, overlapping and unique list-2 pairs occurred equally often. List 3 was created in the same manner. For list-2 pairs, forward-association strength ranged between .014 and .103 and was .051 on average; for list-3 pairs, forward-association strength ranged between .015 and .109 and was .051 on average. All pairs were translated into German.

Design and procedure

The experiment was conducted online via individual meetings using the videotelephony software programme Zoom (Zoom Video Communications). At the start of the experiment, subjects were informed that they should try to remember all word pairs they encountered during the experiment, and that all study material would be tested later. They were *not* explicitly told that they might be exposed to competing information. The experiment consisted of three different phases: an acquisition phase, a distractor phase and a final test phase (see Figure 1). During the acquisition phase, subjects first studied the target list (list 1) and then two subsequent nontarget lists (list 2 and list 3). Before studying each of the three lists, a slide appeared on the computer screen informing subjects that they would now be presented with a particular list. All three lists consisted of 40 stimulus-response pairs. For list 1, half of the stimulus words were presented together with their response words (e.g., plate – *fork*) for 10 s and thus could be studied immediately (study-only condition), while for the other half of list-1 pairs, at first only the stimulus word was presented for 5 s and participants had to guess the corresponding response word (e.g., plate – *?*; pretest condition). Afterwards, the intact word pair was presented for 5 s (e.g., plate – *fork*). Next, all list-2 pairs were shown intact with a presentation rate of 5 s per pair and, finally, all list-3 pairs were shown intact with a presentation rate of 5 s per pair. Recent work showed that in RI tasks, relatively long study times per interpolated pair (10 s) can make it more probable that subjects are reminded of the list-1 response when the interpolated material is studied (Garlitch & Wahlheim, 2020; Negley et al., 2018), thereby preventing RI buildup and promoting later recall of the list-1 response. Since our goal was to induce RI as effectively as possible, we used shorter study times of only 5 s per pair for our lists 2 and 3.

For half of both the study-only and the pretest pairs of list 1, list-2 and list-3 pairs shared the stimulus word but were paired with a new response word (e.g., plate – *fork* in list 1; plate – *glasses* in list 2; plate – *bowl* in list 3; overlapping pairs). For the remaining half of both the study-only and pretest pairs of list 1, the list-2 and 3 pairs consisted of completely new stimulus-response pairs (e.g., plate – *fork* in list 1; atom – *cell* in list 2; territory – *domain* in list 3; unique pairs). Experiment 1 thus applied a 2 × 2 design with the within-subjects factors of TYPE OF PRACTICE (study-only vs. pretest) and TYPE OF PAIRS (unique vs. overlapping). Word pairs of all lists were presented in randomised order for each participant.

Following the acquisition phase, participants played an online spot-the-difference game (https://www.smithsonianmag.com/games/spot-difference-180968040/) for 5 min, before the final test was administered. On this test, memory of all three study lists was assessed. List 1 was always tested first and lists 2 and 3 were tested subsequently. Since we were primarily interested in the

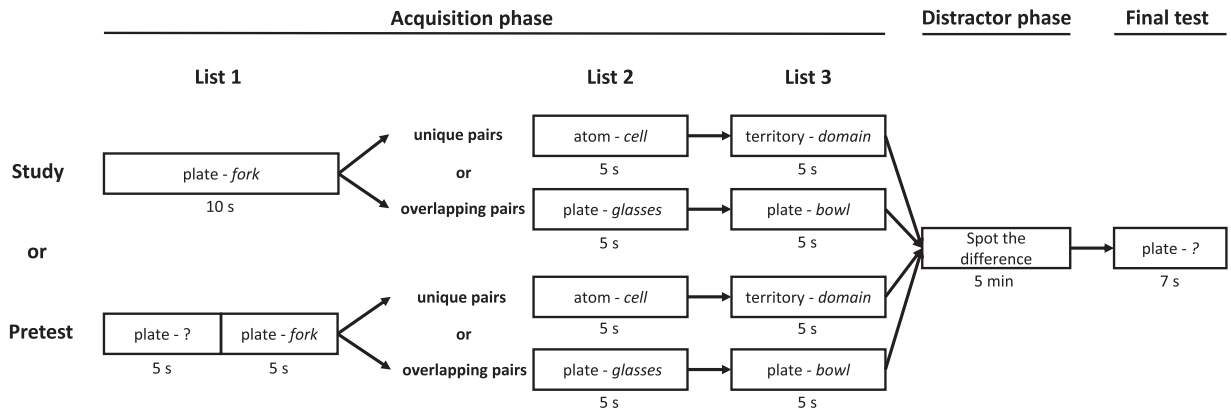

Figure 1. Procedure of Experiment 1. During the initial acquisition phase, half of the 40 list-1 stimulus words were presented together with their response words for 10 s (study condition), while the remaining half of the 40 list-1 stimulus words were pretested for 5 s before corrective feedback was provided for another 5 s (pretest condition). Subjects subsequently studied lists 2 and 3 which both consisted of 40 intact stimulus-response pairs that were each presented for 5 s. Critically, for half of both the 20 study and 20 pretest pairs of list 1, list-2 and list-3 pairs shared the stimulus word but were paired with a new response word (overlapping pairs), while for the remaining half, the list-2 and list-3 pairs were completely new (unique pairs). Following the acquisition phase, participants played a spot-the-difference game for 5 min as a distractor task, before a final cued-recall test on list 1 (the target list) was administered. On this test, the stimulus words of each list-1 pair were presented for 7 s in randomised order and participants were asked to name the corresponding list-1 response. List-2 and list-3 responses were always tested afterwards using the same procedure.

effects of interpolated learning on initially studied material, only list-1 results will be reported in detail in the Results section. For each list, the stimulus words of all 40 pairs were presented for 7 s each in randomised order and participants were asked to name the response word of the tested list (e.g., christmas – ?). Responses during pretesting and the final test were always provided verbally to the experimenter. A response that was similar to a target word was counted as correct only if participants produced a response in its plural form that was originally studied in its singular form, or vice versa (e.g., "books" was counted as correct even when the original target was "book"). No feedback was given during the final test.

Results

Correct recall

On the pretest, participants were able to come up with a guess within the 5-s time frame in 100% of the trials. They correctly guessed 4.5% of the list-1 response words. Those pairs were removed from further analyses, since the effect of erroneous guesses on subsequent memory was of main interest.[1]

Figure 2(a) shows the percentage of correctly recalled list-1 responses on the final test as a function of TYPE OF PRACTICE (study-only vs. pretest) and TYPE OF PAIRS (unique vs. overlapping). A 2 × 2 analysis of variance (ANOVA) of the two factors revealed significant main effects of TYPE OF PRACTICE, $F(1, 39) = 103.54$, $MSE = 79.99$, $p<.001$, $\eta_p^2 = .73$ and TYPE OF PAIRS, $F(1, 39) = 23.26$, $MSE = 79.99$, $p<.001$, $\eta_p^2 = .37$, reflecting overall higher recall rates for pretested pairs than for study-only pairs (82.9% vs. 63.1%) and for unique than overlapping pairs (76.7% vs. 69.4%). These main effects were qualified by a statistically significant interaction between the two factors $F(1, 39) = 9.96$, $MSE = 79.99$, $p = .003$, $\eta_p^2 = .20$.

Pairwise comparisons revealed that, relative to the study-only condition, recall in the pretest condition was improved for unique pairs (84.3% vs. 69.0%), $t(39) = 7.41$, $p<.001$, Cohen's $d = 1.17$ and was improved even more for the overlapping pairs (81.5% vs. 57.3%), $t(39) = 8.98$, $p<.001$, Cohen's $d = 1.42$. Consequently, the size of the pretesting effect increased from 15.3% for unique pairs to 24.3% for overlapping list-1 pairs. Pairwise comparisons further showed that, while in the study-only condition, recall rates were significantly lower for overlapping than unique list-1 pairs (57.3% vs. 69.0%), $t(39) = 5.56$, $p<.001$, Cohen's $d = 0.88$ – resulting in an RI effect of 11.8% – in the pretest condition, there was no reliable difference between overlapping and unique pairs (81.5% vs. 84.3%), $t(39) = 1.40$, $p = .17$, Cohen's $d = 0.22$ – resulting in an RI effect of only 2.8%.

Intrusions overall

Intrusions overall include all commission errors – that is, all overt incorrect responses – which participants produced during the final test of list 1. Figure 2(b) shows the average number of intrusions produced during final testing of list 1 as a function of PRACTICE (study-only vs. pretest) and TYPE OF PAIRS (unique vs. overlapping). A 2 × 2 ANOVA of the two factors revealed main effects of TYPE OF PRACTICE, $F(1, 39) = 64.60$, $MSE = 0.66$, $p<.001$, $\eta_p^2 = .62$ and TYPE OF PAIRS, $F(1, 39) = 13.12$, $MSE = 0.66$, $p<.001$, $\eta_p^2 = .25$, reflecting that, overall, pretesting trials led to a lower number of intrusions than study-only trials (0.8 vs. 2.3 intrusions) and intrusions were lower for unique than overlapping pairs (1.3 vs. 1.9 intrusions). There was also a significant interaction between factors, $F(1, 39) = 5.90$, $MSE = 0.66$, $p = .020$, $\eta_p^2 = .13$. In fact, while pairwise comparisons revealed that initial pretests led to fewer intrusions than initial study for both unique pairs, $t(39) = 5.82$, $p < .001$, Cohen's $d = 0.92$ and

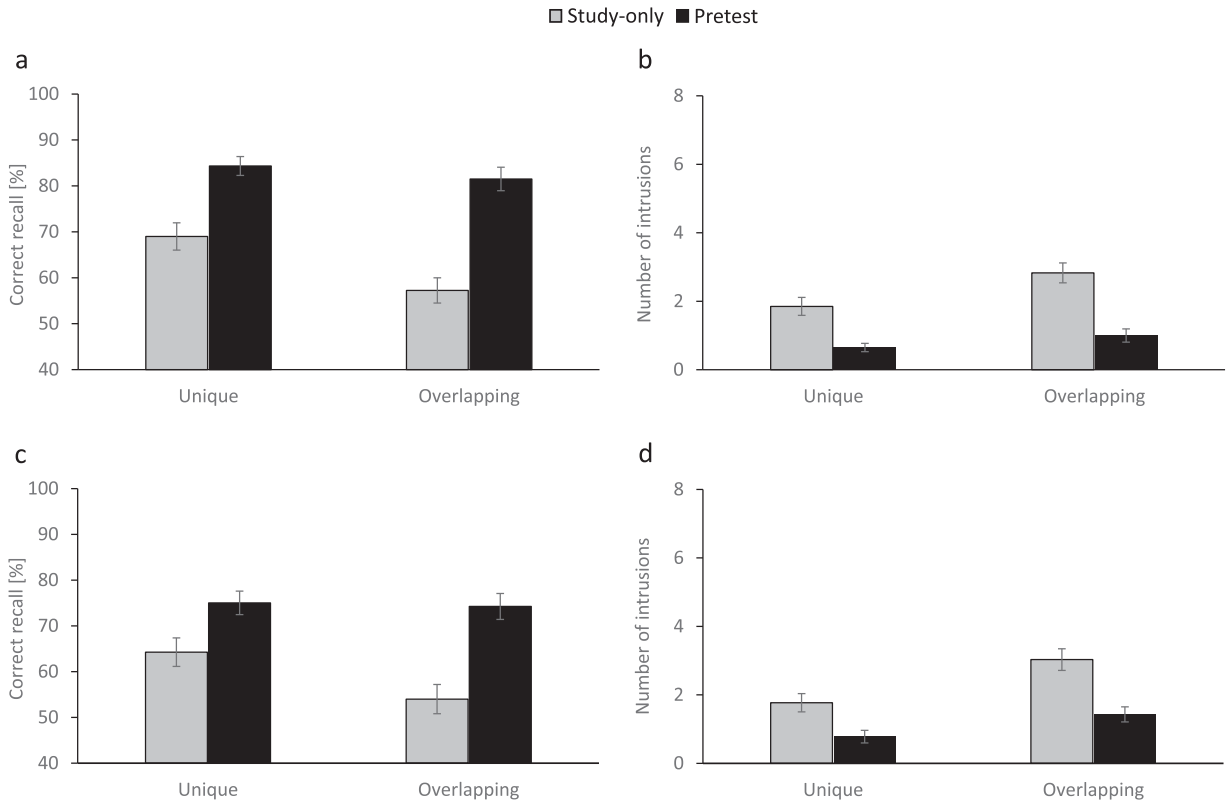

Figure 2. Results of Experiment 1 (a, b) and 2 (c, d). (a) List-1 recall performance on the final test (in %) as a function of TYPE OF PRACTICE (study-only vs. pretest) and TYPE OF PAIRS (unique vs. overlapping). (b) Number of intrusions produced during final testing of list 1 as a function of TYPE OF PRACTICE and TYPE OF PAIRS. (c) List-1 recall performance on the final test (in %) as a function of TYPE OF PRACTICE and TYPE OF PAIRS. (d) Number of intrusions produced during final testing of list 1 as a function of TYPE OF PRACTICE and TYPE OF PAIRS. Error bars reflect standard errors.

overlapping pairs, $t(39) = 7.37$, $p<.001$, Cohen's $d = 1.17$, the magnitude of this benefit of pretesting increased from 1.2 (0.7 vs. 1.9 intrusions) for unique pairs to 1.8 (1.0 vs. 2.8 intrusions) for overlapping pairs.

Interlist intrusions

Interlist intrusions refer to commission errors made during the final test of list 1 which were list-2 or list-3 responses, thus reflecting a subset of the overall intrusions. Arguably, this intrusion measure provides more direct insight into whether pretesting can reduce RI effects since RI is defined as impaired recall of originally studied material due to competition from subsequently studied material. A 2 × 2 ANOVA with the factors TYPE OF PRACTICE and TYPE OF PAIRS revealed main effects of TYPE OF PRACTICE, $F(1, 39) = 13.07$, $MSE = 0.55$, $p<.001$, $\eta_p^2 = .25$ and TYPE OF PAIRS, $F(1, 39) = 38.41$, $MSE = 0.55$, $p<.001$, $\eta_p^2 = .50$, on number of interlist intrusions, reflecting that, overall, pretesting trials led to a lower number of intrusions than study-only trials (0.3 vs. 0.6 intrusions) and intrusions were lower for unique than overlapping pairs (0.1 vs. 0.8 intrusions), aligning with the expectation of higher levels of RI for overlapping than unique pairs. There was no significant interaction between factors, $F(1, 39) = 3.68$, $MSE = 0.55$, $p = .062$, $\eta_p^2 = .09$.

Discussion

The results of Experiment 1 provide a first demonstration that pretesting can induce release from RI. In particular, recall of study-only list-1 responses was found to be impaired in overlapping pairs relative to unique pairs, thus reflecting typical RI (e.g., Greeno, 1964), but recall of pretested list-1 responses did not differ reliably between overlapping and unique pairs, suggesting release from RI. Furthermore, while in unique pairs, recall of pretested responses was larger than of study-only responses, which reflects the typical pretesting effect (e.g., Kornell et al., 2009), the size of the effect further increased in overlapping pairs. In particular, the increased pretesting effect emerged because recall of study-only, but not pretested, list-1 responses decreased in the presence of RI. The results also showed that, in unique list-1 pairs, number of overall intrusions was reduced for pretested responses, relative to study-only responses, which replicates prior work (e.g., Grimaldi & Karpicke, 2012). Again, this beneficial effect of pretesting increased in overlapping pairs. When only interlist intrusions were taken into account, this benefit of pretesting was not statistically reliable. This is not too surprising, since interlist intrusions were fairly rare even for overlapping study-only pairs, leaving not much room for major reductions in interlist intrusions for overlapping pretest pairs. Experiment 1 thus showed

that, in the presence of competing interpolated study material, initial pretesting can both enhance correct recall and reduce number of overall intrusions at the time of final testing.

The goal of Experiment 2 was to conceptually replicate the findings of Experiment 1. Experiment 2 differed in two aspects from Experiment 1. First, different study material was applied in Experiment 2 than in Experiment 1. Second, although procedural details were largely identical to Experiment 1, all list-2 and list-3 pairs were presented twice in Experiment 2 to increase the lists' degree of learning. The aim was to examine whether pretesting can still shield the study material from competing material when, via a higher degree of learning, a potentially higher level of RI was induced by the interpolated lists.

Experiment 2

Method

Participants

Following the power analysis reported in Experiment 1, 40 students (mean age = 25.1 years) took part in Experiment 2. 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.

Material, design and procedure

In Experiment 2, we again used weakly associated stimulus-response pairs but drew completely new pairs from the Nelson et al. (1998) norms. The forward-association strength of the pairs ranged from .012 to .082 and was .034 on average for list-1 pairs, ranged from .010 to .095 and was .033 on average for list-2 pairs, and ranged from .010 to .107 and was .033 on average for list-3 pairs. Furthermore, all procedural details were identical to Experiment 1 with the sole exception that all word pairs of lists 2 and 3 received one extra study cycle which immediately followed the initial presentation of a list's pairs. Prior to this extra cycle, subjects were informed that all pairs they had just seen would be repeated in a new order. During this second study cycle, all pairs were presented again in a new random order for 5 s each.

Results

Correct recall

On the pretest, participants were able to come up with a guess within the 5-s time frame in 100% of the trials. They correctly guessed 2.5% of the response words. Like in Experiment 1, all pairs that were correctly guessed on the pretest were removed from further analyses.

Figure 2(c) shows the percentage of correctly recalled list-1 responses on the final test as a function of TYPE OF PRACTICE (study-only vs. pretest) and TYPE OF PAIRS (unique vs. overlapping). A 2 × 2 ANOVA of the two factors revealed main effects of TYPE OF PRACTICE, $F(1, 39) = 40.01$, $MSE = 102.82$, $p<.001$, $\eta_p^2 = .51$ and TYPE OF PAIRS, $F(1, 39) = 9.16$, $MSE = 102.82$, $p = .004$, $\eta_p^2 = .19$, reflecting higher overall recall rates for pretested than for study-only pairs (74.6% vs. 59.1%) and for unique than for overlapping pairs (69.6% vs. 64.1%). More important, there was also a significant interaction between factors, $F(1, 39) = 8.73$, $MSE = 102.82$, $p = .005$, $\eta_p^2 = .18$.

Pairwise comparisons revealed that, relative to the study-only condition, recall in the pretest condition was improved for unique pairs (75.0% vs. 64.3%), $t(39) = 3.40$, $p = .002$, Cohen's $d = 0.54$ and even more for the overlapping pairs (74.3% vs. 54.0%), $t(39) = 7.58$, $p < .001$, Cohen's $d = 1.20$. Consequently, the size of the pretesting effect increased from 10.8% for unique list-1 pairs to 20.3% for overlapping list-1 pairs. Pairwise comparisons further showed that, while in the study-only condition, recall rates were significantly lower for overlapping than unique list-1 pairs (54.0% vs. 64.3%), $t(39) = 4.11$, $p<.001$, Cohen's $d = 0.65$ – resulting in an RI effect of 10.3% – in the pretest condition, there was no reliable difference between overlapping and unique list-1 pairs (74.3% vs. 75.0%), $t(39) < 1$ – resulting in an RI effect of only 0.8%.

Intrusions overall

Figure 2(d) shows the average number of intrusions produced during final testing of list 1 as a function of TYPE OF PRACTICE and TYPE OF PAIRS. A 2 × 2 ANOVA of the two factors revealed main effects of TYPE OF PRACTICE, $F(1, 39) = 33.84$, $MSE = 1.31$, $p<.001$, $\eta_p^2 = .47$ and TYPE OF PAIRS, $F(1, 39) = 28.50$, $MSE = 1.31$, $p<.001$, $\eta_p^2 = .42$, reflecting that, overall, pretesting trials led to a lower number of intrusions than study-only trials (1.1 vs. 2.2 intrusions) and intrusions were lower for unique than overlapping pairs (1.3 vs. 2.2 intrusions). There was no significant interaction between factors, $F(1, 39) = 2.76$, $MSE = 1.31$, $p = .11$, $\eta_p^2 = .07$.

Interlist intrusions

The same ANOVA revealed main effects of TYPE OF PRACTICE, $F(1, 39) = 20.86$, $MSE = 0.52$, $p<.001$, $\eta_p^2 = .35$ and TYPE OF PAIRS, $F(1, 39) = 68.38$, $MSE = 0.52$, $p<.001$, $\eta_p^2 = .64$, on number of interlist intrusions, reflecting that, overall, pretesting trials led to a lower number of interlist intrusions than study-only trials (0.3 vs. 1.0 intrusions) and intrusions were lower for unique than overlapping pairs (0.1 vs. 1.2 intrusions), aligning with the expectation of higher levels of RI for overlapping than unique pairs. There was also a significant interaction between factors, $F(1, 39) = 23.13$, $MSE = 0.52$, $p<.001$, $\eta_p^2 = .37$. In fact, while pairwise comparisons revealed that initial pretests did not lead to statistically fewer interlist intrusions than initial study for overlapping pairs, $t(39) = 1.28$, $p = .21$, Cohen's $d = 0.50$, a reliable difference arose for overlapping pairs, $t(39) = 4.88$, $p < .001$, Cohen's $d = 1.56$, with the magnitude of this benefit of pretesting increasing from 0.1 (0.1

vs. 0.2 intrusions) for unique pairs to 1.2 (0.6 vs. 1.8 intrusions) for overlapping pairs.

Effectiveness of the degree-of-learning manipulation

One aim of Experiment 2 was to increase the interpolated lists' degree of learning relative to Experiment 1 and thus to potentially increase the amount of RI. To obtain a rough estimate of whether studying lists 2 and 3 twice in Experiment 2 really increased those lists' recall performance on the final test relative to Experiment 1, a 2 × 2 ANOVA with the between-subjects factor of EXPERIMENT (Experiment 1 vs. Experiment 2) and the within-subjects factor of LIST (list 2 vs. list 3) was applied. To keep the analysis simple, we did *not* take into account the factors of TYPE OF PRACTICE and TYPE OF PAIRS for this analysis. ANOVA revealed a main effect of EXPERIMENT, $F(1, 78) = 8.20$, $MSE = 673.48$, $p<.001$, $\eta_p^2 = .10$, reflecting higher overall recall rates for list-2 and list-3 pairs in Experiment 2 than Experiment 1 (50.0% vs. 38.3%), indicating that the degree-of-learning manipulation was successful. ANOVA further showed a significant main effect of LIST, $F(1, 78) = 25.80$, $MSE = 71.95$, $p<.001$, $\eta_p^2 = .25$, reflecting overall superior recall for list-2 than list-3 items (47.5% vs. 40.7%). There was no significant interaction between factors, $F(1, 78) < 1$.[2]

Discussion

Like Experiment 1, the results of Experiment 2 demonstrate that pretesting can induce release from RI. In particular, recall of study-only list-1 responses was again decreased for overlapping, relative to unique, pairs, reflecting typical buildup of RI, but recall of pretested list-1 responses did not differ between overlapping and unique pairs, suggesting release from RI. While we again found a typical pretesting effect for list-1 responses in the absence of competing information, the size of the pretesting effect increased further in the presence of competing information – once again because recall of study-only, but not pretested, list-1 responses decreased in the presence of RI. Similar to Experiment 1, number of overall and interlist intrusions was on average lower for pretested than study-only list-1 pairs but, unlike in Experiment 1, this benefit of pretesting increased statistically only for interlist – and not for overall – intrusions in the presence of overlapping pairs relative to unique pairs.

Presenting lists 2 and 3 twice during acquisition increased the lists' degree of learning, but did not produce higher levels of RI in the study-only (control) condition than in Experiment 1. While in Experiment 1, the difference in recall performance between unique and overlapping pairs was 11.8% in the study-only condition, in Experiment 2, the difference was 10.3%. This finding disagrees with the results from some classic interference studies that found a higher degree of learning of the interpolated study material to increase forgetting of the originally studied target material (e.g., Barnes & Underwood, 1959; Briggs, 1957). However, these studies typically applied the so-called anticipation method during acquisition of the interpolated lists which may be more effective at inducing higher levels of interference than mere restudy of the interpolated lists. In this method, participants perform posttests during acquisition of the interpolated nontarget list(s) on which they are shown the stimulus word of each pair and have to recall the response word before they are shown the intact pair. The method thus involves a posttest with feedback that occurs after the stimulus-response pair has already been studied. We decided not to use the anticipation method, since it would have led to a task involving both pretesting and posttesting, thus making it impossible to establish to what degree pretesting alone can lead to RI reduction. In contrast to the studies using the anticipation method, other studies in the literature employed restudy or longer presentation times to increase degree of learning without the anticipation method, and found no effect of degree of learning on amount of RI (Bäuml, 1996; see also DaPolito, 1966). Our results are consistent with this line of work.

General discussion

The results of the present two experiments demonstrate that pretesting can protect the tested material from RI. While, on the final test, recall of an originally learned target list of word pairs (list 1) was found to be impaired by the subsequent study of overlapping lists of word pairs (lists 2 and 3) when list-1 learning merely involved study of the pairs, no such impairment arose when list-1 learning involved a pretest of the pairs. Consequently, the size of the pretesting effect was found to be larger in the presence of overlapping than in the presence of unique material.[3] Our results suggest a parallel to prior work examining the effects of posttesting on RI, which found that taking a posttest on initially studied material can protect the tested material from the detrimental effects of subsequently studied, overlapping material (Halamish & Bjork, 2011; Potts & Shanks, 2012). Tests therefore appear to immunize against RI, regardless of whether they are administered before or after the initial target material has been studied.

The present finding that pretesting can shield the tested material from RI is consistent with an extension of the elaboration account of the pretesting effect. If pretesting creates a more elaborate memory trace linking the stimulus and response words than does mere study of a stimulus-response pair (e.g., Huelser & Metcalfe, 2012), then it may become more probable that participants are reminded of a given first-list pair when they study overlapping second-list or third-list pairs (e.g., Wahlheim, 2015). When memory of pretested list-1 responses is later tested, there may be less RI from competing list-2 and list-3 responses, since participants tend to be more aware that those responses were studied later during acquisition. Therefore, recall of pretested pairs at the

time of final testing should be less susceptible to RI than recall of studied pairs, which is exactly what the present results show.

Episodic context accounts of the pretesting effect may also explain our findings (Metcalfe & Huelser, 2020; Overoye et al., 2021). These accounts assume that during encoding of information, participants also store details about the temporal context that is present when they are exposed to the information (Estes, 1955; Mensink & Raaijmakers, 1988). Since memory retrieval is sometimes assumed to boost contextual drift (e.g., Jang & Huber, 2008; Jonker et al., 2013), pretested items may be associated with an enriched and more distinct temporal context than study items. Alternatively, or, additionally, pretesting may also improve the binding of the items to the mental context more so than studying the items (e.g., Hopper & Huber, 2018). Critically, in both cases, the probability that one is reminded of a list-1 response word during later encoding of overlapping information may thus be increased for initially pretested, relative to initially studied, items. Analogous to semantic elaboration, an enriched episodic context may thus also promote the creation of an integrated memory trace including all responses and their temporal occurence during acquisition. In principle, both semantic and episodic elaboration may therefore contribute to a pretesting-induced protection from RI.

At first glance, the present finding that pretesting can shield against RI-induced forgetting also seems to align with an extension of attentional accounts of the prestesting effect (e.g., Potts & Shanks, 2014; Zawadzka & Hanczakowski, 2019). Attentional accounts explain the basic pretesting effect by assuming that pretesting enhances attentional encoding of the response word once it is shown at the end of a pretest trial, which should boost recall of the response word on a later final test. To account for the present results, one may assume that pretesting increases the probability that reminding occurs during encoding of overlapping, interpolated lists, since the original response is encoded more thoroughly, and thus, may be more easily reactivated and integrated. Critically, however, attentional accounts typically assume that pretesting enhances primarily attentive encoding of the response word itself, but does not strengthen the association between stimulus and response word (Potts et al., 2019; Seabrooke et al., 2021). If so, pretesting should not increase the probability of reminding when, during study of overlapping lists, stimulus words from the earlier target list are re-encountered. Consequently, attentional accounts appear more difficult to reconcile with the present findings.

Since the typical posttesting and pretesting tasks share certain features, it could be argued that the prior work showing that posttesting can reduce RI effects (Halamish & Bjork, 2011; Potts & Shanks, 2012) already implies that pretesting should be able to reduce RI effects as well. Such conclusion would have been premature, however. Indeed, one major difference between posttesting and pretesting tasks is that, typically, the number of overt errors produced during pretesting are substantially higher than what is observed during posttesting. Averaged across the present two experiments, subjects produced incorrect guesses on 96.5% of all trials on the list-1 pretest, which is within the typical range for this type of pretesting task (e.g., Kornell et al., 2009). By contrast, Halamish and Bjork (2011) reported that 82% of responses were correct on the posttest, meaning that, at most, 18% of responses were wrong. Since, by design, pretest trials thus create much more errors than posttest trials, and the difference in error rates may well influence the cognitive processes that are involved in the two effects, the finding that pretesting reduces RI should not be considered a simple consequence of the parallel posttest finding. Rather, the two findings may indicate that the two test effects share some cognitive processes – for instance those that influence interference susceptibility – but differ in other cognitive processes – for instance, those that enable the integration of the stimulus and cue items either in the presence (pretesting) or absence (posttesting) of erroneous guesses. Identifying the two classes of processes is a high priority for future research.

Together with recent work reporting that pretesting can prevent forgetting when the delay prior to final testing is prolonged (Kliegl et al., in press), the present results on reduced RI suggest a high potential of pretesting for educational practice. Indeed, two of the major challenges which students often encounter in real-life learning situations are that information needs to be reproduced after prolonged periods of time and that it has to be reproduced in the presence of competing information. Pretesting thus may be helpful as a learning tool to manage both of these challenges, but further research is required. While Kliegl et al. (in press) have already shown that pretesting can prevent forgetting over time when prose material is used as study material, future work is required to evaluate whether pretesting can also prevent interference-induced forgetting with more educationally relevant study material. The effectiveness with which pretests can reduce RI for more complex material may, however, depend on whether the pretest questions are likely to elicit plausible responses. Indeed, if participants are unable to provide even remote guesses to pretest questions, the pretested material may not be protected from RI. One option would be to use interrelated prose passages as study material, for instance, with each passage containing certain types of economic and demographic facts about different countries (e.g., passage 1 contains facts about Spain; passage 2 contains facts about South Africa; passage 3 contains facts about Switzerland), and participants receiving a pretest on some of the facts of passage 1 (e.g., "What is Spain's main export product?"; correct answer: "Cars") before studying passage 1 and, subsequently, passages 2 and 3. If the present results with word pairs generalised to such prose passages, chances

would further increase that pretesting can reduce RI also in real-life settings.

To conclude, the results of the present two experiments show that pretesting can shield the tested material from the detrimental effects of induced RI. This finding is consistent with extensions of both elaboration and episodic context accounts of the pretesting effect, but appear less consistent with attentional accounts of the effect. From an applied perspective, the current findings suggest that using pretesting as an educational tool may be useful in learning contexts in which students are exposed to competing information.

Notes

1. For both experiments, we also checked whether the pattern of results was affected when only overlapping list-1 pairs were included into the analyses of recall data for which *neither* the list-1, list-2, or list-3 response words were guessed during pretesting. Using this more restrictive exclusion criterion did not change the general pattern of results, with the critical TYPE OF PRACTICE ×TYPE OF PAIRS interaction persisting in both experiments, all ps $\leq$ .006.
2. Because the mean forward-association strength of the interpolated word pairs in Experiment 2 was slightly lower than in Experiment 1, it is possible that the extent of the degree-of-learning manipulation is somewhat underestimated.
3. For unique pairs, the observed pretesting effect essentially replicates the pretesting benefit observed in typical (single-list) pretesting tasks since the study of further unique pairs in lists 2 and 3 should not lead to considerable RI effects when the unique list-1 pairs are later tested.

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 (KL 2967/2-1).

Open practices statement

The data and materials for all experiments are available at https://osf.io/csr38/, and none of the experiments was preregistered.

References

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97–105. https://doi.org/10.1037/h0047507

Bäuml, K.-H. (1996). Revisiting an old issue: Retroactive interference as a function of the degree of original and interpolated learning. *Psychonomic Bulletin & Review*, *3*(3), 380–384. https://doi.org/10.3758/BF03210765.

Bäuml, K.-H. T., & Kliegl, O. (2017). Retrieval-induced remembering and forgetting. In J. T. Wixted, & J. H. Byrne (Eds.), *Cognitive psychology of memory, Vol. 2 of learning and memory: A comprehensive reference* (2nd ed., pp. 27–51). Academic Press. http://doi.org/10.1016/B978-0-12-809324-5.21048-1

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

Briggs, G. E. (1957). Retroactive inhibition as a function of the degree of original and interpolated learning. *Journal of Experimental Psychology*, *53*(1), 60–67. https://doi.org/10.1037/h0047458

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133. https://doi.org/10.1037/a0019902

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. (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*(6), 1547–1552. https://doi.org/10.1037/a0024140

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. (2012). Trial-and-error learning improves source memory among young and older adults. *Psychology and Aging*, *27*(2), 429–439. https://doi.org/10.1037/a0025115

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*(9), 1281–1290. https://doi.org/10.1080/09658211.2018.1464189

DaPolito, F. J. (1966). *Proactive effects with independent retrieval of competing responses* [Unpublished doctoral dissertation]. Indiana University.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*(3), 145–154. https://doi.org/10.1037/h0048509.

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

Garlitch, S. M., & Wahlheim, C. N. (2020). The role of reminding in retroactive effects of memory for older and younger adults. *Psychology and Aging*, *35*(5), 697–709. https://doi.org/10.1037/pag0000427

Greeno, J. G. (1964). Paired-associate learning with massed and distributed repetitions of items. *Journal of Experimental Psychology*, *67*(3), 286–295. http://dx.doi.org/10.1037/h0044977

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

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

Hopper, W. J., & Huber, D. E. (2018). Learning to recall: Examining recall latencies to test an intra-item learning theory of testing effects. *Journal of Memory and Language*, *102*, 1–15. https://doi.org/10.1016/j.jml.2018.04.005

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514–527. https://doi.org/10.3758/s13421-011-0167-z

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition*, *41*(5), 625–637. https://doi.org/10.3758/s13421-013-0298-5

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *34*(1), 112–127. https://doi.org/10.1037/0278-7393.34.1.112

Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review*, *120*(4), 852–872. https://doi.org/10.1037/a0034246

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). Elsevier. http://doi.org/10.1016/B978-0-12-809324-5.21055-9

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

Kliegl, O., & Bäuml, K.-H. T. (2021). The mechanisms underlying interference and inhibition: A review of current behavioral and neuroimaging research. *Brain Sciences*, *11*(9), 1246. https://doi.org/10.3390/brainsci11091246

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

Mensink, G. J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434–455. https://doi.org/10.1037/0033-295X.95.4.434

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

Müller, G. E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie*, *1*, 1–300.

Negley, J. H., Kelley, C. M., & Jacoby, L. L. (2018). The importance of time to think back: The role of reminding in retroactive effects of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1352–1364. http://dx.doi.org/10.1037/xlm0000512

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/

Overoye, A. L., James, K. K., & Storm, B. C. (2021). A little can go a long way: Giving learners some context can enhance the benefits of pretesting. *Memory*, *29*(9), 1206–1215. https://doi.org/10.1080/09658211.2021.1974048

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, *1*(1), 19–40. https://doi.org/10.3758/BF03198064.

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

Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics. 2010.09.003

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

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

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, *43*(1), 27–38. https://doi.org/10.3758/s13421-014-0455-5

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

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