Journal of Experimental Psychology: Applied

Repeated Guessing Attempts During Acquisition Can Promote Subsequent Recall Performance

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

Online First Publication, August 10, 2023. https://dx.doi.org/10.1037/xap0000493

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

https://doi.org/10.1037/xap0000493

Repeated Guessing Attempts During Acquisition Can Promote Subsequent Recall Performance

Oliver Kliegl, Johannes Bartl, and Karl-Heinz T. Bäuml
Department of Experimental Psychology, Regensburg University

Taking a pretest before to-be-learned material is studied can improve long-term retention of the material relative to material that was initially only studied. Using weakly associated word pairs (Experiments 1 and 3), Swahili–German word pairs (Experiment 2), and prose passages (Experiment 4) as study material, the present study examined whether this pretesting effect is modulated in size when pretests are repeatedly administered during acquisition. All four experiments consistently showed the typical pretesting effect, with enhanced recall after a single guessing attempt relative to the study-only baseline. Critically, the pretesting effect increased in size when multiple guessing attempts were made during acquisition, regardless of whether the duration of the pretesting phase increased with the number of guesses (Experiments 1, 2, and 4) or was held constant (Experiment 3). The results of Experiment 4 also indicate that neither a single guessing attempt nor multiple guessing attempts easily induce the transfer of learning to previously studied but untested information. Together, the findings demonstrate that additional guesses can promote access to the pretested target material on the final test, suggesting that in educational contexts, extensive pretesting during acquisition may serve as an effective learning strategy.

Public Significance Statement

The pretesting effect refers to the finding that taking a test on material that has not yet been studied can improve long-term retention of the material, even when the initial guess is wrong. In the present study, we demonstrate that asking participants to come up with multiple unique guesses during acquisition can even increase the pretesting effect in size, regardless of whether subjects studied weakly associated word pairs, such as knife–*pond*, translations of Swahili words, such as mashua–*boat*, or prose passages covering topics such as the country of Brazil. The findings suggest that when applying pretesting as a teaching tool in the classroom, it may be particularly beneficial if, in response to a prequestion, instructors have students come up with multiple unique guesses.

Keywords: testing effect, pretesting effect, repeated guessing, elaboration

Retrieval from memory changes the information retrieved. A particularly striking demonstration of this phenomenon can be observed in studies on the testing effect, in which subjects who are asked to retrieve previously studied material from memory perform better on a later final test compared to subjects who reread the material (e.g., Roediger & Karpicke, 2006). A multitude of studies have shown that the testing effect is general and robust, arising across a wide range of study materials and ability levels in both lab-based studies (see Karpicke, 2017; Rowland, 2014, for reviews) and educational settings (see Dunlosky et al., 2013; Yang et al., 2021, for reviews).

Previous research on the pretesting effect has illustrated that performing a pretest on some information followed by an opportunity to study the information can also promote later retention of that material (see Kornell & Vaughn, 2016, for a review). In a typical task, subjects are either asked to study a weakly associated cue–target word pair (e.g., plate–*fork*) or they are initially shown the cue item for a few seconds and then have to provide a guess as to the associated target item (plate–*?*) before the cue–target pairing is shown intact. On a later final test, in which the target item has to be recalled from the cue item (plate–*?*), recall performance is typically improved for target

This work was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft) awarded to Oliver Kliegl (KL 2967/2-1). The study design, hypotheses, and analytic plan were not preregistered. The authors have no conflicts of interest to disclose.

All study materials and data have been made publicly available on the Open Science Framework and can be accessed at https://osf.io/q48by/.

Oliver Kliegl played a lead role in conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, and writing–original draft and an equal role in data curation, supervision, and writing–review and editing. Johannes Bartl played a supporting role in visualization and writing–review and editing and an equal role in data curation and investigation. Karl-Heinz T. Bäuml played a lead role in resources, a supporting role in conceptualization, funding acquisition, investigation, methodology, project administration, and writing–original draft, and an equal role in supervision and writing–review and editing.

Correspondence concerning this article should be addressed to Oliver Kliegl, Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany. Email: Oliver.Kliegl@psychologie.uni-regensburg.de

items that were initially pretested compared to target items that were initially studied (e.g., Grimaldi & Karpicke, 2012).

In the pretesting task, subjects can only guess the target item. As a result, they barely ever succeed at retrieving the correct response. Still, even if those very few trials on which correct guesses were made are removed from further analysis, a pretesting effect typically arises, thus suggesting that pretesting trials on which failed retrieval attempts are followed by corrective feedback can be more helpful at promoting memory than simple study trials. Crucially, pretesting even promotes memory if, on study trials, the cue–target pairs are presented as long—for example, 10 s—as the summed guess and presentation duration on pretest trials—for example, 5 s guess duration and 5 s presentation duration (e.g., Kornell et al., 2009). While not yet investigated as extensively as the testing effect, the pretesting effect has also been observed in both lab-based studies and educational settings across a variety of study materials, including weak associates, videos, trivia questions, and foreign language learning (see Chan et al., 2018; Kornell & Vaughn, 2016, for reviews).

The demonstration that making errors during learning can promote later retention of the target material may seem surprising, given that a body of research on errorless learning has shown that making errors during learning can actually impair memory of the target material (e.g., Baddeley & Wilson, 1994; Skinner, 1957). Furthermore, a variety of memory models assume that the probability of accessing target information decreases as more information is associated with a given cue, an assumption referred to as cue overload (e.g., Raaijmakers & Shiffrin, 1981; Watkins & Watkins, 1975). Because such additional information may also include erroneous guesses made during pretesting, the guesses could compete with the retrieval of target information on the final test, thereby impairing recall of the target information.

In contrast, the elaboration hypothesis of the pretesting effect suggests that errors made during acquisition can support later retrieval of the target item (Carpenter, 2009, 2011; Huelser & Metcalfe, 2012). The proposal is that when, during an initial pretest, subjects are shown a cue item (plate–*?*) and have to guess the target item, memory representations related to the cue item (this may involve items like *table*, *meal*, *knife*) are activated and become associated with the cue–target pair once feedback has been provided. On the final test, these memory representations, which also include the erroneous guess made during the pretest, may serve as semantic mediators and aid retrieval of the target item, thus creating the pretesting effect. The finding that pretests not only improve later recall of the target item but also slow down its retrieval (Huelser & Metcalfe, 2012) is in line with the elaboration account of the pretesting effect since attempting to produce an initially pretested item may first lead to the retrieval of mediating information, which may slow recall of the target item.

The cue-overload principle and the elaboration hypothesis of the pretesting effect lead to very different expectations on how repeated initial pretests might influence recall of target information. On the basis of the elaboration hypothesis, the expectation may arise that repeated initial pretests benefit the retention of the study material. For instance, if subjects were asked to give two or three guesses—instead of just one guess—this might lead to an even more extensive activation of information related to the cue item. Moreover, the incorrect guesses might serve as additional semantic mediators that facilitate access to the target item on the final test, and thus lead to a more pronounced pretesting effect. Alternatively, on the basis of the cue-overload principle, adding more erroneous information to the cue as a result of repeated pretesting might increase competition when subjects attempt to access the target information at test, thus decreasing the probability of correct recall. Depending on the theoretical view, a more extensive pretest might therefore enlarge or reduce the pretesting effect.

The question of whether repeated pretesting can promote retention of the material to be learned is of high relevance for pedagogical practice. Previous work investigating the role of pretesting as a teaching tool has already shown that the pretesting effect can occur when educationally relevant study materials are used in both the laboratory (Hilaire & Carpenter, 2020; Kliegl et al., in press; Overoye et al., 2021; Toftness et al., 2018) and the classroom (Carpenter et al., 2018) for retention intervals of up to 7 days. These findings raise the question of whether instructors should spend even more time and effort on pretesting. For instance, before addressing a particular topic—for example, the types of mineral resources certain countries possess—a teacher might ask their students not only to give one possible answer to a pretest question about the topic (e.g., "What mineral resources does Sweden have?"), but to give several answers before the topic is covered. Knowing whether or not such additional investment in a pretest can ultimately pay off is of critical importance for application.

The Present Study

Four experiments were conducted to investigate whether repeated retrieval attempts during acquisition can increase the size of the pretesting effect. In Experiments 1–3, a typical pretesting effect task was employed, in which subjects either studied intact cue–target word pairs or were first only presented with the cue item of each pair and asked to guess the target item before being presented with the intact cue–target pair (e.g., Kornell et al., 2009). Unlike in a typical task, however, participants were asked on some of the initial pretesting trials immediately after making the first guess to either provide one further guess (Experiment 1) or two further guesses (Experiments 1–3). On a later retention test, subjects' memory of all word pairs was assessed by providing a pair's cue item and asking subjects to produce the correct target item. In Experiments 1 and 3, weakly associated word pairs were used as study material, while in Experiment 2, Swahili–German[1] word pairs served as study material. In Experiments 1 and 2, the length of a pretesting trial increased with the number of guesses. In contrast, in Experiment 3, the duration of all pretesting trials was held constant to rule out possible effects of the amount of time spent on guessing the size of the pretesting effect. In Experiment 4, subjects studied three prose passages, with each passage covering a different topic (such as the country of Brazil). While one of the passages was shown in full from the beginning, in the two remaining passages, some keywords were missing, and subjects had to make either one guess or three guesses about the missing information before this information was revealed.

For all four experiments, a reliable pretesting effect was expected following pretest trials involving a single guess attempt compared to study-only trials (e.g., Kornell et al., 2009). It was unclear, however, whether the more effortful pretesting would lead to a more

[1] All subjects participating in Experiments 1–3 spoke German as their first language.

pronounced or less pronounced pretesting effect. While both the elaboration hypothesis and the cue-overload principle assume that repeated pretesting should lead to an enriched memory trace associated with the cue information, the elaboration hypothesis assumes that the additional information can support access to the target information, while the cue-overload principle suggests that the additional information rather constitutes a source of interference and thus should impair access to the target information. A third option would be that the effects of the two (nonexclusive) processes cancel out each other on the final test, leaving the size of the pretesting effect largely unaffected. The results of the present study should provide answers on whether repeated guessing attempts during acquisition are beneficial, detrimental, or largely irrelevant for the size of the pretesting effect.

Experiment 1

Experiment 1 examined the effects of repeated retrieval attempts on the size of the pretesting effect using weakly associated word pairs as study material. To this end, participants either studied the word pairs intact, or they were asked to provide a single guess, two guesses, or three guesses before the correct target item was revealed. German translations of weakly associated word pairs (e.g., frog–*pond*) served as study material. In the study condition, the cue and the target items were presented together for 10 s (e.g., frog–*pond*; study–10 s condition). In the one-guess condition, the cue item was presented alone for 5 s and subjects were asked to guess the target item (e.g., frog–*?* one guess–5 s condition) before the cue and the target items were presented together for 5 s. In the two- and three-guesses conditions, subjects had 10 s or 15 s to make two or three guess attempts in immediate succession (two guesses–10 s and three guesses–15 s conditions) before the cue and target items were presented together for 5 s. We used these rather complex labels for our conditions to highlight the differences between conditions across experiments. Following this acquisition phase, participants were tested on all target items (e.g., frog–*?*). Final test performance was assessed by measuring correct recall and intrusions.

Method

Transparency and Openness

All study materials and data have been made publicly available on the Open Science Framework and can be accessed at https://osf.io/q48by/. The analytic code is not available via Open Science Framework because the analyses reported in this article involve only standard two-factorial analyses of variance (ANOVAs) and planned comparisons, which are already sufficiently characterized in the Result sections of the present Experiments 1–4.

Participants

To determine the sample size in Experiments 1–3, 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, 48 students ($M_{age} = 24.5$ years; 40 females, eight males, 0 diverse) were recruited to participate in Experiment 1. 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.

Material

Sixty weakly associated cue–target pairs (e.g., plate–*fork*, tradition–*christmas*) drawn from the Nelson et al. (1998) norms were used as study material. The forward association strength of each word pair was in the narrow range of .051–.053 and was .052 on average, meaning that subjects in the Nelson et al. study produced a target item approximately 5% of the time as their first associate when provided with a given cue item. All items were translated into German and were divided into four subsets. Each subset consisted of 15 word pairs and was used equally often in each of the experimental conditions.

Design

The experiment used a within-subjects design with four conditions: study–10 s versus one guess–5 s versus two guesses–10 s versus three guesses–15 s. Conditions differed with respect to whether word pairs were presented intact (study–10 s condition), or only the cue item was shown and participants were asked to generate one response (one guess–5 s condition), two responses (two guesses–10 s condition), or three responses (three guesses–15 s condition) before the complete pair was shown.

Procedure

The whole experiment was conducted online via individual meetings using the videotelephony software program Zoom (Zoom Video Communications). Each participant went through three phases: a study phase, a distractor phase, and a final test phase. During the study phase, all 60 word pairs were presented sequentially, but only 15 of the pairs were shown intact from the beginning. On these study–10 s trials, the cue item was presented together with its target item (plate–*fork*) for 10 s and thus could be studied immediately. The remaining 45 word pairs received pretests. Fifteen of those pairs were assigned to one guess–5 s trials, in which, at first, the cue item was presented for 5 s and participants tried to guess the corresponding target item (e.g., tradition–*?*). At the beginning of each trial, a tone was presented for 0.5 s to indicate that a guess attempt should now be made. Another 15 pairs each were assigned to two guesses–10 s and three guesses–15 s trials, on which participants were given either one or two more 5 s periods to make one or two more guesses. The 0.5 s tone was presented at the beginning of each 5-s period to indicate that another guess should be made. It was emphasized that participants should make a unique guess each time. Participants did not know at the beginning of each guess trial how many guessing attempts were required. All three types of guess trials ended with the intact word pairs being presented for 5 s (e.g., tradition–*christmas*). As a result, study–10 s and one guess–5 s trials both lasted 10 s overall, two guesses–10 s trials lasted 15 s overall, and three guesses–15 s trials lasted 20 s overall. For each participant, word pairs were presented in a randomized order. Following the study phase, participants were asked to count backward in steps of seven for 1 min before playing an online spot-the-difference game (https://www.suchbilder.com/fehlerbilder/) for another 4 min. Next, the final test was administered. On this test, the cue items of all 60 word pairs were presented for 10 s each in randomized order and participants were asked to name the corresponding target (e.g., plate–*?*). No feedback was given during the final test.[2]

Results

Initial Pretest

An overall ANOVA of the three pretesting conditions (one guess–5 s, two guesses–10 s, three guesses–15 s) showed a significant effect of condition on guessing performance, $F(2, 94) = 7.05$, $MSE = 22.20$, $p = .001$, $\eta_p^2 = .13$. Planned comparisons showed that subjects produced significantly fewer correct guesses in the one guess–5 s condition, relative to the two guesses–10 s condition (4.6% vs. 7.6%), $t(47) = 3.36$, $p = .002$, Cohen's $d = 0.49$, whereas the difference in correct guesses was not significant between the two guesses–10 s and three guesses–15 s condition (7.6% vs. 7.8%), $t(47) < 1$. Since we were interested in the effects of erroneous guesses on subsequent memory, items that were correctly guessed during the pretest were excluded from further analyses.

Final Test

Correct Recall. Figure 1a shows the percentage of correctly recalled items on the final test for each of the four conditions (study–10 s vs. one guess–5 s vs. two guesses–10 s vs. three guesses–15 s). An overall ANOVA of the four conditions showed a significant effect of condition on correct recall, $F(3, 141) = 112.17$, $MSE = 34.10$, $p < .001$, $\eta_p^2 = .71$. Planned comparisons showed that the difference of 10.2 percentage points between the study–10 s and the one guess–5 s condition was reliable, $t(47) = 7.24$, $p < .001$, Cohen's $d = 1.05$, demonstrating the typical pretesting effect. Furthermore, the difference of 5.6 percentage points between the one guess–5 s and the two guesses–10 s condition was significant, $t(47) = 5.93$, $p < .001$, Cohen's $d = 0.86$, as was the difference of 5.1 percentage points between the two guesses–10 s and the three guesses–15 s condition, $t(47) = 5.04$, $p < .001$, Cohen's $d = 0.73$. The size of the pretesting effect thus increased from 10.2 percentage points to 15.8 percentage points to 20.8 percentage points as the number of initial guesses increased.

Discussion

The results of Experiment 1 provide a first indication that the size of the pretesting effect can increase when repeated retrieval attempts are made during acquisition. While a reliable pretesting effect arose in response to a single guessing attempt—as reflected in an improved recall performance for the one guess–5 s condition relative to the study–10 s condition—the size of the pretesting effect roughly doubled from the one guess–5 s to the three guesses–15 s conditions. The aim of Experiment 2 was to address the generalizability of these findings.

Experiment 2

Experiment 2 investigated whether the beneficial effects of repeated pretesting still emerge when more educationally relevant study material is applied. To this end, 60 Swahili–German word pairs were used as study material, half of which each participant studied intact and half of which received at least a single pretest prior to study. In particular, 15 pairs received a single pretest (one guess–5 s condition), while another 15 pairs received three pretests (three guesses–15 s condition). Unlike in Experiment 1, a second study condition—the study–20 s condition—was included in Experiment 2, for which we ensured that the total trial duration matched the duration of three guesses–15 s trial.[3] As a result, 15 of the 30 pairs that were presented intact were assigned to study–10 s trials in which pairs were shown intact for 10 s each, thus matching the overall duration of one guess–5 s trials. The remaining 15 pairs were assigned to study–20 s trials in which pairs were shown intact for 20 s each, thus matching the overall duration of three guesses–15 s trials.

Method

Participants

Following Experiment 1, 48 students ($M_{age} = 24.4$ years; 39 females, nine males, 0 diverse) were recruited to take 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 reward.

Material

The study material consisted of 60 Swahili–English word pairs (e.g., tumbili–*monkey*) taken from the Nelson and Dunlosky (1994) norms, for which the target items were translated into German. All pairs included had a frequency of less than 100 occurrences per million words according to the Hyperspace Analogue to Language frequency scale (Balota et al., 2007). Analogous to Experiment 1, all word pairs were divided into four subsets with 15 word pairs each, with each subset being used equally often in each of the four experimental conditions.

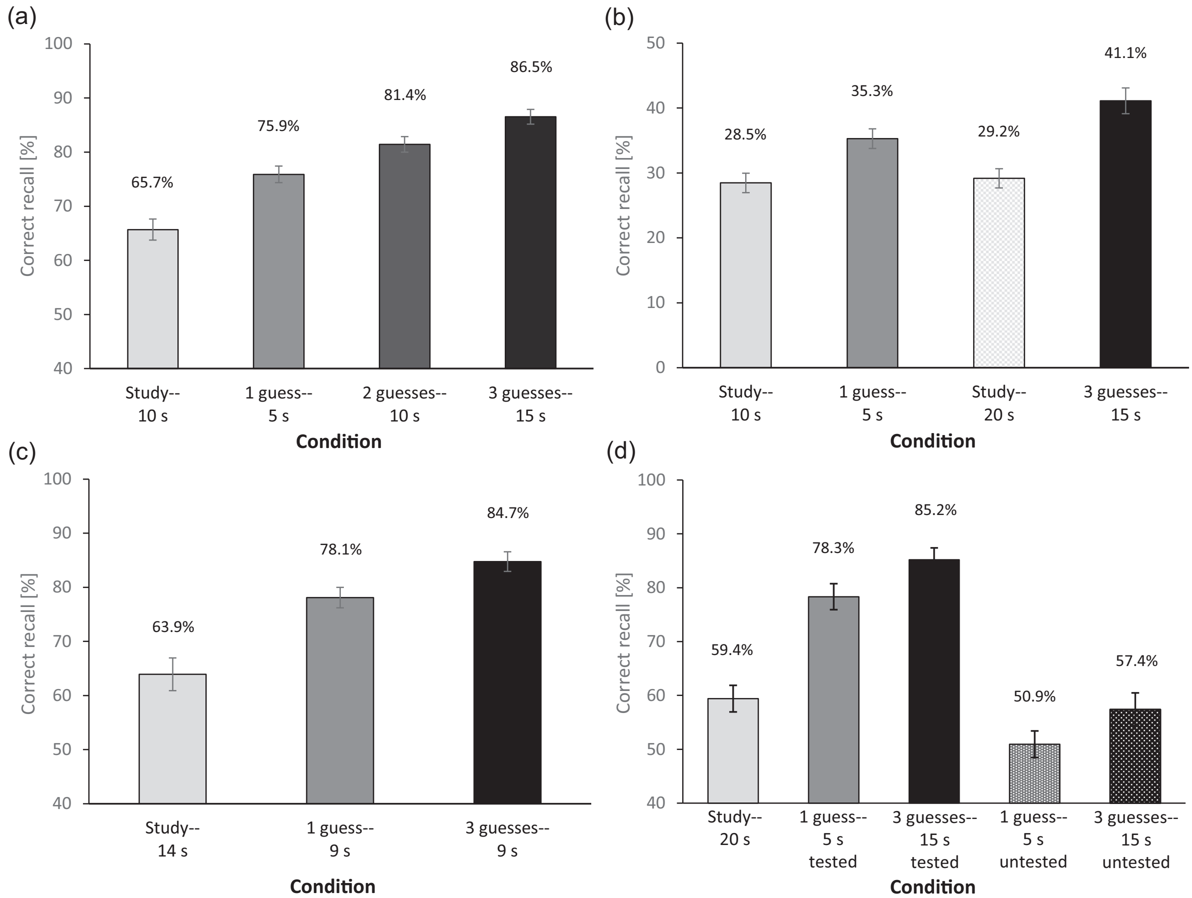
Design and Procedure

Procedural details were identical to Experiment 1, with the exception that the two guesses–10 s condition of Experiment 1 was replaced by a study–20 s condition, which served as a baseline to pretested items for which three guesses were made. As a result, Experiment 2 had a 2 × 2 design with the within-subjects factors of type of practice (study vs. pretest) and amount of practice (low vs. high). Half of the word pairs were presented intact and thus could be studied immediately (study condition), while for the other half of the pairs, participants were asked to guess the target items before the complete pair was shown (pretest condition). On pretest trials, participants either had one guess attempt (low amount of practice) or three guess attempts (high amount of practice). To equate trial duration for the pretest and study conditions, participants were

[2] All study materials and data have been made publicly available on the Open Science Framework and can be accessed at https://osf.io/q48by/.

[3] Since the study condition used in Experiment 1 already provided ample time (i.e., 10 s) for studying a given word pair, we deemed it unlikely that an even longer study duration would considerably affect recall performance. However, to be completely sure of the effects of prolonged study trials on recall, it was nevertheless decided to add such a baseline condition in Experiment 2.

Figure 1
Results of Experiments 1–4



Note. (a) Results of Experiment 1. Recall performance on the final test (in %) for each of the four conditions (study–10 s vs. one guess–5 s vs. two guesses–10 s vs. three guesses–15 s). All guessing trials ended with an additional 5-s period in which the intact cue–target pair was presented. (b) Results of Experiment 2. Recall performance on the final test (in %) for each of the four conditions (study–10 s vs. one guess–5 s vs. study–20 s vs. three guesses–15 s). All guessing trials ended with an additional 5-s period in which the intact cue–target pair was presented. (c) Results of Experiment 3. Recall performance on the final test (in %) for each of the three conditions (study–14 s vs. one guess–9 s vs. three guesses–9 s). All guessing trials ended with an additional 5-s period in which the intact cue-target pair was presented. (d) Results of Experiment 4. Recall performance on the final test (in %) for the study–20 s condition and, separately for tested and untested sentences, for the one guess–5 s and the three guesses–15 s conditions. All guessing trials ended with an additional 5-s period in which the missing word was presented. Error bars reflect the standard error of the mean.

shown the complete word pair for 10 s longer—that is, for 20 s overall—on study trials in the high amount of practice than the low amount of practice condition.

Results

Initial Pretest

During initial pretesting, none of the target items were guessed correctly. Therefore, all items were included in the further analyses.

Final Test

Correct Recall. Figure 1b shows the percentage of correctly recalled target items on the final test for each of the four conditions (study–10 s vs. one guess–5 s vs. study–20 s vs. three guesses–15 s). A 2 × 2 ANOVA with the factors of type of practice (study vs. pretest) and amount of practice (low vs. high) revealed main effects of type of practice, $F(1, 47) = 62.99$, $MSE = 40.30$, $p < .001$, $\eta_p^2 = .57$, and amount of practice, $F(1, 47) = 13.78$, $MSE = 40.30$, $p = .001$, $\eta_p^2 = .23$, reflecting higher overall recall rates for pretested items than for studied items (38.2% vs. 28.8%) and in the high-amount than the low-amount condition (35.1% vs. 31.9%). More importantly, there was a significant interaction between the two factors, $F(1, 47) = 7.86$, $MSE = 40.30$, $p = .007$, $\eta_p^2 = .14$. Planned comparisons revealed that relative to their corresponding study (control) condition, recall of pretested items was improved for the low-amount condition, $t(47) = 4.83$, $p < .001$, Cohen's $d = 0.70$, and, to a larger extent, also for the high-amount condition, $t(47) = 7.58$, $p < .001$, Cohen's $d = 1.10$. Consistently, the magnitude of the pretesting effect increased from 6.8 percentage points (35.3% vs. 28.5%) to 11.9 percentage points (41.1% vs. 29.2%) from the low-amount to the high-amount condition.

Discussion

The results of Experiment 2 generalize the findings of Experiment 1 by showing that repeated retrieval attempts can still increase the size of the pretesting effect when more educationally relevant study materials are used. Indeed, when participants were asked to study Swahili–German word pairs and provide three guesses during pretesting, the resulting pretesting effect was almost twice as large compared to when only one guess was required.

Experiment 3

The experimental tasks used in Experiments 1 and 2 involve a natural confounding since, in both experiments, the duration of a pretest trial increased with the number of guesses that had to be generated. It thus remains unclear whether it is the number of guesses per se that was responsible for the observed increase in the pretesting effect or rather the amount of time spent on the task. Experiment 3 addressed the issue and examined whether acquisition trials involving repeated guessing attempts are still more beneficial for subsequent recall performance than acquisition trials requiring only a single guessing attempt when the duration of the two types of trials is held constant.

Like in Experiment 1, German translations of weakly associated word pairs (e.g., frog–*pond*) served as study material. Like in Experiment 2, participants were either asked during pretesting to come up with one guess or were asked to come up with three guesses before the correct target item was revealed. Unlike in the previous two experiments, however, participants had 9 s to produce a guess during trials involving a single guess but only 3 s per guess during trials involving three guesses, which resulted in the same total trial duration for the two types of pretesting trials. If the number of guessing attempts was the critical factor increasing the size of the pretesting effect in Experiments 1 and 2, then the size of the pretesting effect should still be more pronounced for trials involving three guesses than for trials involving a single guess in the present experiment. In contrast, if the prolonged amount of time spent on repeated guessing trials was the more relevant factor for the increased pretesting effect that was observed in the two preceding experiments, then the size of the pretesting effect should be largely unaffected by the number of guessing attempts.

Method

Participants

Following Experiment 1, 48 students (M_{age} = 24.9 years; 38 females, 10 males, 0 diverse) were recruited to take 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 reward.

Material

The same study material as in Experiment 1 was applied. The assignment of word pairs to the experimental conditions followed the same scheme as in Experiments 1 and 2.

Design and Procedure

Experimental details of Experiment 3 were similar to Experiment 1, with two critical exceptions. First, the experiment used a within-subjects design with three (instead of four) conditions: study–14 s versus one guess–9 s versus three guesses–9 s. In study–14 s trials, the word pairs were shown intact for 14 s; in one guess–9 s trials, participants had 9 s to produce a guess before the complete pair was shown for 5 s; in three guesses–9 s trials, participants had 3 s to produce each of their guesses before the word pair was shown for 5 s. Consequently, both one guess–9 s and three guesses–9 s trials lasted 14 s overall. Second, one guess–9 s and three guesses–9 s trials were presented in a blocked format during the study phase: For half of the participants, the study phase started with a first block in which they were shown 15 of the cue items and had 9 s to come up with a single guess attempt for each cue before the cue–target pairs were shown intact for 5 s. The 15 trials were randomly intermixed with 15 study trials in which the cue–target pairs were shown intact for 14 s. This first block was immediately followed by a second block in which another 15 cue items were shown and participants were given 3 s for each of three guessing attempts before the cue–target pairs were shown intact for 5 s. The 15 trials were again randomly intermixed with 15 study trials in which the cue–target pairs were shown intact for 14 s. For the remaining half of the participants, the order of blocks was reversed. Participants were informed about the number of required responses on all guessing trials prior to the start of each block.

Results

Initial Pretest

A pairwise comparison showed that subjects produced significantly fewer correct guesses in the one guess–9 s condition relative to the three guesses–9 s condition (4.4% vs. 9.7%), $t(47) = 5.55$, $p < .001$, Cohen's $d = 0.80$. Since we were interested in the effects of erroneous guesses on subsequent memory, items that were correctly guessed during the pretest were excluded from further analyses.[4]

Final Test

Correct Recall. Because mean recall performance for the study–14 s trials of the two blocks did not differ, $t(47) < 1$, Cohen's $d = 0.01$, all study–14 s trials were pooled for further analyses. Figure 1c shows the percentage of correctly recalled items on the final test for each of the three conditions (study–14 s vs. one guess–9 s vs. three guesses–9 s). An overall ANOVA of the three conditions showed a significant effect of condition on correct recall, $F(2, 94) = 44.10$, $MSE = 123.08$,

[4] A direct comparison between the number of correct guesses produced in Experiments 1 and 3 showed no reliable difference for both the one-guess condition (4.6% vs. 4.4%), $t(94) < 1$, Cohen's $d = 0.03$, and the three-guesses condition (7.8% vs. 9.7%), $t(94) = 1.47$, $p = .14$, Cohen's $d = 0.30$, suggesting that the amount of time available per guessing attempt played no critical role for the success rate in guessing.

$p < .001$, $\eta_p^2 = .48$. Planned comparisons showed that the difference of 14.2 percentage points between the study–14 s and the one guess–9 s condition was reliable, $t(47) = 5.53$, $p < .001$, Cohen's $d = 0.80$, demonstrating the typical pretesting effect. Furthermore, the difference of 6.7 percentage points between the one guess–9 s and the three guesses–9 s condition was significant, $t(47) = 3.82$, $p < .001$, Cohen's $d = 0.55$, with the size of the pretesting effect thus increasing from 14.2 percentage points to 20.8 percentage points from the one guess–9 s to the three guesses–9 s condition.

Discussion

Like the present Experiments 1 and 2, the results of Experiment 3 showed that the size of the pretesting effect can increase when multiple guessing attempts are made during acquisition. Critically, this pattern arose even though total trial duration was held constant for the one-guess and three-guesses trials, suggesting that the increase in size of the pretesting effect observed in the current Experiments 1 and 2 with additional guessing attempts should not have been caused by prolonged trial duration. Rather, the findings of Experiment 3 suggest a critical contribution of number of guesses to the magnitude of this testing effect.

Experiment 4

Experiment 4 had two main objectives. The first aim was to investigate whether repeated guessing attempts during acquisition can still promote memory when more educationally relevant study materials (i.e., prose passages) are used. The second aim was to examine whether repeated guessing attempts would also promote the memory of previously studied but untested information. Examining whether pretesting can afford also transfer to untested material appears important from an applied perspective since teachers often avoid revealing exam questions during instruction (e.g., Wooldridge et al., 2014). While prior work suggests that typically no transfer to untested material will be observed—at least with only a single guessing attempt (e.g., James & Storm, 2019; Richland et al., 2009)—in these studies, all pretest questions were answered in a blocked format before the material was studied. However, pretesting-induced elaboration processes may be relatively short-lived since the magnitude of the pretesting effect is typically more pronounced when the correct-answer feedback follows immediately after the pretest compared to when the feedback is delayed (e.g., Grimaldi & Karpicke, 2012). Thus, if, during acquisition, the presentation of untested material occurred in close temporal proximity to the pretests, such elaboration processes might more easily spill over to the untested material. In particular, the pretest-induced activation of mediators may still be present when a subsequent, untested sentence is shown a few seconds later. Since at least a subset of these mediators may show some degree of semantic relatedness to the information in the untested sentence, the information in the untested sentence may become better integrated into memory—and easier to recall on the final test—than in the absence of pretesting. Experiment 4 examined this possibility.

All subjects studied three prose passages. One of the passages was presented in the study-only format. The sentences were displayed on a computer screen one after the other and could be studied immediately (e.g., "Woodrow Wilson was the 28th president of the United States."). The two remaining passages were presented in the one-guess and three-guesses formats. In both cases, half of the sentences were first displayed in a fill-in-the-blank format, and participants had to make one guess or three guesses about the missing information (e.g., "Woodrow Wilson was the ____ [insert number] president of the United States.") before the full sentence was shown. Critically, these fill-in-the-blank sentences alternated with complete sentences (i.e., the untested information). In the later final test, subjects were tested on both the initially pretested and the untested information. On the basis of the results of Experiments 1–3, we expected that the size of the pretesting effect would increase with number of initial guesses, thus generalizing the results with paired associates to more complex prose passages. If the closer temporal proximity of pretests to the presentation of untested material is a crucial factor to induce the transfer of learning to untested material, then a single guess and, in particular, three guesses made during acquisition might also improve memory of untested material on the final test compared to the study-only baseline.

Method

Participants

Following Experiment 1, 54 students ($M_{\text{age}} = 25.6$ years; 39 females, 15 males, 0 diverse) were recruited to take part in Experiment 4. Due to counterbalancing, six more participants were tested than in the three previous experiments. 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 reward.

Material

The study material consisted of three encyclopedically prepared German text passages about former U.S. president Woodrow Wilson (233 words), the country of Brazil (266 words), and Tapirs (252 words). The texts contained facts taken from Wikipedia (https://en.wikipedia.org/wiki/Woodrow Wilson; https://en.wikipedia.org/ wiki/Brazil; https://en.wikipedia.org/wiki/Tapir). Each text consisted of 12 sentences and was used equally often in the study-only condition, the one-guess condition, and the three-guesses condition. For each sentence, a final test question was constructed based on a fact contained in the sentence. The questions were phrased as short open-answer questions (e.g., "What was the first name of Woodrow Wilson's mother?" or "When did Woodrow Wilson's presidency begin?"). These 12 questions were divided into two subgroups, consisting of six questions each (Subgroups A and B). For half of the participants, the information that was asked for in the questions of Subgroup A was missing in the initially presented fill-in-the-blank sentences, while for the remaining half of participants, the information that was asked for in the questions of Subgroup B was missing in the initial fill-in-the-blank sentences. On the final test, subjects were asked to provide an answer to all 12 questions.

Design and Procedure

Like Experiments 1–3, Experiment 4 was conducted via Zoom. The experiment used a within-subject design with three conditions: study–20 s, one guess–5 s, three guesses–15 s. In each condition, subjects first studied one of the three texts and then completed a final test on that passage. During the study phase of the study–20 s condition, the 12 sentences of a passage were shown sequentially on a computer screen at a rate of 20 s per sentence. During the study phase of the one guess–5 s condition, the 12 sentences of a passage were also shown sequentially at a rate of 20 s per sentence, but only half of the sentences were shown completely from the beginning. For the remaining half of the sentences, one word was missing, and subjects were prompted 10 s after the onset of the sentence via a tone presented for 0.5 s to guess the missing word. Subjects had 5 s to make their guess before the missing word appeared for 5 s. The study phase of the three guesses–15 s condition was similar to the study phase of the one guess—5 s condition with the sole exception that three unique guesses had to be made—and subjects were given 5 s for each guess—before the missing word was revealed. In both the one guess—5 s and three guesses–15 s conditions, tested and untested sentences were always presented alternatingly. Following the study phase, participants counted backward in steps of seven for 1 min and then played the online spot-the-difference game for another 4 min, before completing a final test on the immediately preceding text. On this test, participants received the 12 questions on the text, with one question per sentence. Subjects were given 8 s to answer a question. Following the final test, participants were asked to play Tetris (https://www.geo.de/geolino/spiele/13349-rtkl-online spiel-tetris) for 3 min before the next text passage was presented.

Results

Initial Pretest

A planned comparison showed that the difference in correct guesses was not significant between the one guess–5 s and three guesses–15 s conditions (3.1% vs. 5.2%), $t(54) = 1.63$, $p = .11$, Cohen's $d = 0.22$. Since we were interested in the effects of erroneous guesses on subsequent memory, items that were correctly guessed during the pretest were excluded from further analyses.

Final Test

Figure 1d shows the percentage of correctly answered questions on the final test for the study–20 s condition and, separately for tested and untested sentences, the percentage of correctly answered questions for the one guess–5 s and the three guesses–15 s conditions. In the first step, an overall ANOVA comparing correct recall of the study–20 s condition with correct recall of the one guess–5 s and three guesses–15 s conditions for pretested material showed a reliable effect, $F(2, 106) = 42.21$, $MSE = 227.94$, $p < .001$, $\eta_p^2 = .44$. Planned comparisons revealed that the difference of 18.9 percentage points between the study–20 s and the one guess–5 s condition was reliable, $t(53) = 5.94$, $p < .001$, Cohen's $d = 0.81$, demonstrating the typical pretesting effect, as was the difference of 25.8 percentage points between the study–20 s and three guesses–15 s conditions, $t(53) = 9.15$, $p < .001$, Cohen's $d = 1.25$.

In a second step, an overall ANOVA comparing correct recall of the study-only condition with correct recall of the one guess–5 s and three guesses–15 s conditions for untested material also showed a reliable effect, $F(2, 106) = 3.091$, $MSE = 227.94$, $p = .05$, $\eta_p^2 = .06$. Planned comparisons revealed that while the difference of 8.5 percentage points between the study–20 s and the one guess–5 s condition was reliable, $t(53) = 2.44$, $p = .018$, Cohen's $d = 0.33$, the difference of 2.0 percentage points between the study–20 s and three guesses–15 s conditions was not statistically significant, $t(53) < 1$, Cohen's $d = 0.08$.

In a third and final step, we conducted a 2 × 2 ANOVA with the factors of type of practice (one guess–5 s vs. three guesses–15 s) and type of material (tested vs. untested) to examine whether the increase in recall rates from the one guess–5 s condition to the three guesses–15 s condition differed for tested and untested material. ANOVA showed main effects of type of practice, $F(1, 53) = 8.91$, $MSE = 337.70$, $p = .004$, $\eta_p^2 = .14$, and type of material, $F(1, 53) = 125.89$, $MSE = 337.70$, $p < .001$, $\eta_p^2 = .70$, reflecting higher overall recall rates after three guesses than a single guess (71.3% vs. 64.4%) and for tested than untested material (81.8% vs. 54.17%). Critically, there was no interaction between factors, $F(1, 53) < 1$, suggesting that three guesses, relative to a single guess, resulted in a similar boost in recall performance for tested and untested material.

Discussion

The results of Experiment 4 show a reliable benefit of pretesting for pretested information following a single guessing attempt, thus replicating prior work demonstrating the pretesting effect for prose passages (e.g., James & Storm, 2019; Richland et al., 2009). Critically, the size of the pretesting effect increased when three guesses were made, thus generalizing the findings of the present Experiments 1–3 with paired associates to more complex study material. Furthermore, the current results suggest that pretesting did not improve later recall of untested information, both when a single guessing attempt and when three guessing attempts were made, which fits with prior work reporting no beneficial effect of a single guessing attempt on recall of previously studied but untested material (e.g., James & Storm, 2019; Richland et al., 2009). Recall in the one-guess condition was even inferior to recall in the study-only condition. Still, recall of untested material showed a similar increase from the one-guess to the three-guesses conditions as recall of tested material did, thus leading to comparable recall levels for the untested material for the three-guesses and study-only conditions (see General Discussion).

Additional Analysis

Most studies examining the pretesting effect primarily focus on correct recall performance on the final test as the dependent variable, and only a few studies also analyze commission errors, that is, intrusions (e.g., Grimaldi & Karpicke, 2012; Kliegl et al., in press). We also chose not to provide a detailed analysis of final test intrusions. Rather, Table 1 provides a descriptive overview of intrusions produced in Experiments 1–4 as a function of the three study formats that were common to all four experiments, that is, the study condition, the one-guess condition, and the three-guesses condition.[5] These intrusions were subjected to an overall analysis to get a broad estimate as to the effects of (repeated) initial guessing on the number of intrusions. For this analysis, we pooled the intrusion rates—that is, the number of intrusions divided by the overall number of study items in a condition—of all four experiments for the study condition, the one-guess condition, and the three-guesses condition.

[5] Since Experiment 3 had two study conditions, only the study–10 s condition was included in the table.

Table 1
Mean Number of Intrusions (Standard Error of the Mean) Produced During the Final Test as a Function of Study Format for Experiments 1–4

Experiment	Study	One guess	Three guesses
Experiment 1	2.77 (0.27)	1.38 (0.17)	0.83 (0.13)
Experiment 2	3.06 (0.26)	1.63 (0.22)	1.58 (0.20)
Experiment 3	2.85 (0.38)	1.46 (0.21)	0.88 (0.17)
Experiment 4	2.78 (0.21)	0.78 (0.12)	0.39 (0.09)

An overall ANOVA of the three conditions revealed a significant effect of condition on the intrusion rate, $F(2, 394) = 81.81$, $MSE = 75.39$, $p < .001$, $\eta_p^2 = .29$. Planned comparisons showed that the difference of 7.8 percentage points (18.8% vs. 11.0%) between the study-only and the one-guess conditions was reliable, $t(197) = 7.58$, $p < .001$, Cohen's $d = 0.54$, as was the difference of 3.8 percentage points (11.0% vs. 7.2%) between the one-guess and the three-guesses conditions, $t(197) = 4.73$, $p < .001$, Cohen's $d = 0.34$. Results thus replicate prior work showing that a single initial guessing attempt can reduce number of intrusions (e.g., Grimaldi & Karpicke, 2012; Kliegl et al., in press) and extend the prior work by demonstrating that this benefit of pretesting increases when three guesses, instead of one guess, have to be made. Together with the correct recall data, the intrusion data suggest that repeated guessing attempts not only promote later recall of tested material but also prevent intrusion errors. We also analyzed "pretest" intrusions, which occur only when participants produce an erroneous initial guess on the final test. Naturally, this analysis only included the one-guess and three-guesses conditions of Experiments 1–4 and did not reveal a statistical difference (3.1% vs. 2.5%), $F(1, 197) = 1.56$, $MSE = 17.32$, $p = .214$, $\eta_p^2 = .01$. At least numerically, the pattern of results is, however, similar to the overall analysis above, and the missing statistical difference between conditions may be due to a floor effect problem. The finding may also appear remarkable, as the number of *potential* pretest intrusions is three times higher in the three-guesses than in the one-guess condition.

General Discussion

The present study provides the first demonstration that the benefits of pretesting can be enhanced when repeated retrieval attempts are made during acquisition. Employing weakly associated word pairs (Experiments 1 and 3), Swahili–German pairs (Experiment 2), and prose passages (Experiment 4), it was found that the size of the pretesting effect can increase as the number of initial guessing attempts increases, with the magnitude of the pretesting effect enlarging between roughly 35% (Experiment 4) and roughly 100% (Experiments 1 and 2) in response to three guessing attempts, relative to a single guessing attempt. In contrast, initial guessing attempts, be they one or three erroneous guesses, do not seem to boost later recall of untested information, suggesting that pretesting cannot readily induce transfer effects on previously studied but untested information.

The present finding that repeated guessing attempts can increase the size of the pretesting effect for pretested material is consistent with the elaboration account of the pretesting effect. The account assumes that pretesting leads to a more comprehensive activation of memory representations related to the cue item, which, on the final test, may be used as semantic mediators through which the target items are retrieved. On the basis of this account, the prediction arises that repeated guessing attempts can lead to a more elaborate memory trace than a single guessing attempt, with the erroneous guesses serving as additional mediators through which access to the target item is facilitated on the final test. Consequently, the size of the pretesting effect should be more pronounced following multiple guessing attempts relative to a single guessing attempt, which is exactly what the results of Experiments 1–4 demonstrate.

In contrast, the findings do not align with the cue-overload principle. According to this view, multiple guessing attempts during acquisition should impede access to the target material on the final test since the erroneous guesses constitute a source of competition. The principle would thus predict that, on the final test, recall of target items is worse after multiple guesses than after a single guess (or no guess). However, the present findings clearly suggest that multiple erroneous guesses do not impede, but rather support, access to the target information.

If elaboration really mediated the pretesting effect for the pretested information in the present experiments and elaboration spilled over to untested information, then recall of untested information should also be higher after three guessing attempts than one single guessing attempt. The results of Experiment 4 indeed show such pattern, which first of all supports the elaboration view. However, the same line of reasoning also suggests that recall of untested information after a single guessing attempt—and the more so after three attempts—should be higher than in the study-only condition, which is not what Experiment 4 suggests. Moreover, the result of a detrimental effect of pretesting with one guessing attempt disagrees with other prior work, in which neutral effects of pretesting on recall of untested information were found (e.g., James & Storm, 2019). A possible explanation for the observed detrimental effect with one guessing attempt might be that pretest sentences drew so much attention that subjects largely neglected, focusing on subsequently presented untested sentences. If so, the detrimental effect might disappear if subjects were instructed to equally focus on study-only sentences, and a beneficial effect of pretesting arise at least after three guessing attempts if the instruction was successful. Future work may examine this possibility, which might provide interesting insights on whether elaboration mediated recall of both the pretested and the untested information.

While the present results for tested material are consistent with an elaboration explanation of the pretesting effect, they are also consistent with attentional and episodic context accounts of the effect. Attentional accounts assume that pretesting boosts attentional encoding of the subsequent feedback (i.e., when the target item is revealed), thus resulting in enhanced recall of the target item on the final test (e.g., Potts & Shanks, 2014). On the basis of this view, providing additional guesses might further increase subjects' curiosity, thus recruiting even more attentional resources to encode the target information and increasing the size of the pretesting effect. Episodic context accounts of the pretesting effect (Metcalfe & Huelser, 2020) are based on the assumption that when we encode information, features of the temporal context that is present when we encounter that information are stored (Estes, 1955; Mensink & Raaijmakers, 1988). Pretested items could therefore be associated with an enriched temporal context consisting of an integrated representation of the pretest context and the study context, while the study items should be associated with the study context only. Repeated pretesting may thus be particularly beneficial since each guessing attempt may elicit a (slightly) new temporal context, resulting in a greater variety of contextual elements associated with the pretested items, which may facilitate retrieval of the items in a subsequent retention test.

Like the elaboration account, the attentional and episodic context accounts have trouble explaining the detrimental effect of pretesting on recall of untested information after one single guessing attempt. Unlike the elaboration account, the two accounts can also not easily explain why three guessing attempts improve recall of untested information relative to a single guessing attempt. Indeed, attentional accounts assume that pretest-induced enhancements in attentional processing should be limited to the tested information, irrespective of whether a single guessing attempt or multiple guessing attempts were made. Similarly, episodic context accounts assume that pretest-induced enrichments in mental context should pertain only to tested information, regardless of the number of guessing attempts. When taking both recall of pretested and untested information into account, one may thus prefer the elaboration account over the other two accounts, although future conceptual enrichments of the attentional and episodic context accounts may overcome this shortcoming.

The results of the present study demonstrate that repeated guessing attempts during acquisition can increase the size of the pretesting effect regardless of whether trial duration increases with number of guesses (Experiments 1, 2, 4) or is held constant (Experiment 3). This finding does not exclude a potential role of trial duration for the size of the pretesting effect, in particular, as, at least numerically, a comparison of the sizes of the pretesting effects observed in Experiments 1 and 3 suggests a somewhat smaller effect in Experiment 3 than in Experiment 1. However, the results of a previous study indicate that simply increasing the duration of a guessing trial does not increase the magnitude of the pretesting effect (Kornell & Vaughn, 2016). In this study, participants were either asked to guess the correct answer to trivia questions (Experiments 1–3) or they were shown three words and asked to name a fourth word related to those three (*lip* would be the solution to *reading*, *service*, and *stick*; Experiment 4) before the correct answer was shown. While in all four experiments, participants had either 5, 10, or 30 s to make a guess, there was no manipulation of the number of retrieval attempts. Kornell and Vaughn found a pretesting effect when subjects were given 5 s to provide a guess relative to the study condition and found the magnitude of the effect to remain largely unchanged when additional time was reserved for pretesting.

While the present findings provide a relatively clear demonstration that repeated pretesting can be beneficial for the pretested information, the results from two earlier studies (partly) deviate from this pattern. In one study, Vaughn and Rawson (2012) found that producing three guesses during pretesting enhanced later recall of the target information relative to producing a single guess, which is consistent with our results. However, no pretesting effect was observed for both the one-guess and the three-guesses conditions, which may have been caused by the fact that the researchers chose to provide delayed feedback during pretesting, a procedure that has repeatedly been shown to prevent reliable pretesting effects (e.g., Grimaldi & Karpicke, 2012). In the other study, James (2022) reported across three experiments no beneficial or detrimental effects of repeated pretesting when trivia questions were used as study material, raising the possibility that the present results with paired associates and prose passages may not generalize to trivia questions. Future research appears thus necessary to examine the potential role of study material for the effectiveness of repeated pretesting in further detail.

From an applied perspective, the current results indicate that using pretests in educational settings may be particularly beneficial when multiple erroneous guesses are made before the correct response is provided. The results of the present Experiment 2 suggest that repeated pretesting can even be beneficial in situations in which the probability of guessing the correct response is (close to) zero. A geography teacher may thus ask students to provide multiple guesses as to the little-known capital of a country (such as the capital of Switzerland; correct answer: "Bern") or a history teacher may want students to come up with several guesses about the years in which a particular U.S. president held office (such as William Howard Taft; correct answer: "1909–1913"). However, to bridge the gap from the present findings to such real-life situations, further work is required that examines the effects of repeated pretesting in real classroom scenarios, using actual course material.

To conclude, the findings of the present study suggest that pretesting can improve retention, and having to generate multiple guesses during initial pretesting can even enhance memory for study material. From a theoretical perspective, this observation aligns with the elaboration account of the pretesting effect, supporting the view that repeated guesses can increase the elaboration of the study material. From an applied perspective, the results suggest that pretesting can still serve as an efficient learning tool when multiple (erroneous) responses are made prior to study, with the benefits of repeated guessing even eclipsing the benefits of a single guessing attempt. A possible limitation of (repeated) pretesting may be that its benefits do not transfer to previously studied but untested information.

References

Baddeley, A., & Wilson, B. A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*, *32*(1), 53–68. https://doi.org/10.1016/0028-3932(94)90068-X

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*, 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), Article 31996. https://doi.org/10.1525/collabra.31996

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

Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied*, *24*(1), 34–42. https://doi.org/10.1037/xap0000145

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

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 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

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

Hilaire, K. J. S., & Carpenter, S. K. (2020). Prequestions enhance learning, but only whentheyareremembered. *Journal of Experimental Psychology: Applied*, *26*(4), 705–716. https://doi.org/10.1037/xap0000296

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

James, K. K. (2022). *Do more guesses lead to more benefits? On the consequences of multiple attempts in the pretesting effect*. University of California, Santa Cruz.

James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested? *Journal of Experimental Psychology: Applied*, *25*(4), 576–587. https://doi.org/10.1037/xap0000231

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. https://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*. https://doi.org/10.1037/mac0000085

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2023, April 21). *Repeated guessing attempts during acquisition can promote subsequent recall performance*. https://osf.io/q48by

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*, 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*, Article 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*. https://w3.usf.edu/FreeAssociation/

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili–English translation equivalents. *Memory*, *2*(3), 325–335. https://doi.org/10.1080/09658219408258951

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

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–677. https://doi.org/10.1037/a0033194

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243–257. https://doi.org/10.1037/a0016496

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

Skinner, B. F. (1957). *Verbal behavior*. Prentice Hall.

Toftness, A. R., Carpenter, S. K., Lauber, S., & Mickes, L. (2018). The limited effects of prequestions on learning from authentic lecture videos. *Journal of Applied Research in Memory and Cognition*, *7*(3), 370–378. https://doi.org/10.1037/xap0000145

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review*, *19*, 899–905. https://doi.org/10.3758/s13423-012-0276-0

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442–452. https://doi.org/10.1037/0278-7393.1.4.442

Wooldridge, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational mate-rials: A cautionary note. *Journal of Applied Research in Memory & Cognition*, *3*, 214–221. https://doi.org/10.1016/j.jarmac.2014.07.001

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Received October 20, 2022
Revision received April 21, 2023
Accepted June 30, 2023 ■