Journal of Applied Research in Memory and Cognition

The Pretesting Effect Comes to Full Fruition After Prolonged Retention Interval

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

Online First Publication, November 28, 2022. https://dx.doi.org/10.1037/mac0000085

AMERICAN PSYCHOLOGICAL ASSOCIATION SARMAC Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681
https://doi.org/10.1037/mac0000085

EMPIRICAL ARTICLE

The Pretesting Effect Comes to Full Fruition After Prolonged Retention Interval

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

Taking a pretest before material is studied can enhance recall of that material on a subsequent final test. In the present research, we examined whether the magnitude of this pretesting effect is modulated by the delay that precedes the final test. Experiment 1 employed paired associates as study material and retention intervals of up to 30 min; Experiment 2 employed an educationally more relevant prose passage as study material and retention intervals of up to one whole week. In both experiments, we examined whether pretesting some of the study material improved recall of the pretested information relative to other material that was not pretested. Results of both experiments replicated the benefit of pretesting for retention of studied material. Strikingly, this pretesting effect increased and roughly doubled with increasing delay. Pretesting could play a significant role in educational settings where information typically needs to be retained over longer periods of time.

General Audience Summary

Testing material to be learned, like paired associates (e.g., plate–*fork*) or a prose passage (e.g., about the big bang theory), shortly after study can be highly beneficial for retention of the information—as demonstrated by the wealth of research on the so-called testing effect. Somewhat counterintuitively, this benefit of testing does not only arise when testing takes places after study but can also emerge when testing precedes study, and individuals are thus forced to guess the (unknown) correct answer—typically with little success. To be of potential use for educational settings, this so-called pretesting effect—like the classic testing effect—should survive prolonged retention intervals between acquisition and final testing and not be restricted to situations in which the final test follows shortly upon acquisition. We addressed the issue in two experiments. The one experiment employed paired associates as study material and retention intervals of up to 30 min; the other experiment employed an educationally more relevant prose passage as study material and retention intervals of up to one whole week. In both experiments, we examined whether pretesting some of the study material improved recall of the pretested information relative to other material that was not pretested. Both experiments replicated the benefit of pretesting for retention of studied information. Strikingly, in both experiments, the pretesting effect roughly doubled from the short to the longest retention interval. These findings demonstrate that the pretesting effect is not a transient effect but can even increase in size as the retention interval between study and final test is prolonged. The pretesting effect thus fulfills a critical precondition to be useful in educational settings, with its typically longer periods of time between acquisition and final testing of to-be-learned material.

Keywords: testing effect, pretesting effect, retention interval, elaboration

Johannes Bartl https://orcid.org/0000-0003-1547-6450
Karl-Heinz T. Bäuml https://orcid.org/0000-0002-4966-9288

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

All dependent variables or measures that were analyzed for this article's target research question have been reported in the Method section. Also, all levels of all independent variables or all predictors and manipulations, whether successful or failed, have been reported in the Method section. No observations were excluded.

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

Further requests for the data or materials can be sent via email to the corresponding author at (oliver.kliegl@ur.de).

Neither of the experiments reported in this article was formally preregistered. However, the study materials employed in the present experiments as well as the data from the single experiments are available on the Open Science Framework at https://osf.io/n7bqz/?view_only=3f2f21b4f70440efabeab53f7aa103cd.

Correspondence concerning this article should be addressed to Oliver Kliegl, Department of Experimental Psychology, Regensburg University, Universitätsstrasse 31, Regensburg 93040, Germany. Email: oliver.kliegl@psychologie.uni-regensburg.de

In recent years, studies on the so-called testing effect have demonstrated that memory tests not only serve to unveil a person's current state of knowledge but also represent learning events in themselves. Indeed, tests of previously studied information often lead to better recall of the information in a later retention test than, for example, passive restudy of this information does (e.g., Roediger & Karpicke, 2006). The testing effect is a robust and general phenomenon that occurs across a wide range of materials, age groups, and ability levels (for reviews, see Karpicke, 2017; Roediger & Butler, 2011). Testing can promote longer term learning even in classroom settings with real course materials (for reviews, see Agarwal et al., 2021; Yang et al., 2021).

In the classic testing effect task, participants first study a set of items before, in the testing (practice) phase, they are asked to retrieve the items from memory. However, a testing effect can also result when this order is reversed, and participants are tested before the actual information is studied. In a typical study demonstrating such a pretesting effect, participants in the (baseline) study condition are presented with a cue item together with a target item (plate–*fork*) for study. In contrast, in the pretest condition, participants first have to guess the target item in the presence of the cue item (plate–*?*) before the correct answer is shown (plate–*fork*). Because subjects are not shown the word pairs prior to the initial pretest and correct guess attempts are not considered in the data analysis, these studies isolate the effects of failed retrieval attempts. Studies using this task generally show that pretesting leads to better recall performance for the target items than simply studying the cue–target pairs. This holds even when the total presentation duration of the individual word pairs in the study condition is as long—for example, 10 s—as the summed guess and presentation duration in the pretest condition—for example, 5 s guess duration + 5 s presentation duration (e.g., Kornell et al., 2009). The pretesting effect has been found over a wide range of study materials, such as weak associates (Grimaldi & Karpicke, 2012), videos (Carpenter & Toftness, 2017; James & Storm, 2019), general knowledge questions (Kornell, 2014), and prose passages (Hilaire & Carpenter, 2020; Richland et al., 2009).

Soon after publication of the first pretesting studies, pretesting was already recommended for use as an instruction tool in educational practice (Pashler et al., 2007). Yet, in order to serve as an effective learning technique in the classroom, it is critical to understand whether and how the pretesting effect depends on length of retention interval between acquisition and final testing of the study material. Prior research has already demonstrated that the pretesting effect can persist with pedagogically relevant study material both in the laboratory (Hilaire & Carpenter, 2020; Overoye et al., 2021; Toftness et al., 2018) and in classroom learning (Carpenter et al., 2018) for retention intervals of up to 1 week. However, one of the features that makes the classic testing effect interesting for the pedagogical context is the finding that the size of the effect not only persists but often becomes even larger with increasing retention interval (for a meta-analysis, see Rowland, 2014). For instance, in their seminal study, Roediger and Karpicke (2006) had participants study a prose passage which was either immediately tested or restudied before a final retention test was conducted after either 5 min, 2 days, or 1 week. While there was even a slight disadvantage for tested, relative to restudied, information when the final test was performed after 5 min, tested information was much better retained than restudied information when the final test was conducted after 2 days or 1 week.

Whether the size of the pretesting effect also increases, or rather decreases, with retention interval is unclear. Based on prior work indicating that source-monitoring accuracy declines as the length of the retention interval increases (e.g., Johnson, 1997; Johnson et al., 1993), one may expect a reduced pretesting effect with delay. Indeed, with longer delays, participants may be less well able to distinguish between their initial (erroneous) guesses and the target item to be retrieved at the time of final testing. Another indication against the conjecture that prolonged retention interval can benefit the pretesting effect comes from a meta-analysis conducted by Chan et al. (2018) that included data from 45 individual experiments and found that the size of the pretesting effect tended to decrease with length of retention interval. However, in all but one of the included studies (i.e., Kornell et al., 2009), length of retention interval varied up to 15 min only, leaving it open whether results would generalize to much longer retention interval. Furthermore, with the exception of a single study (Overoye et al., 2021; see General Discussion section below), the previous studies also refrained from systematically manipulating length of retention interval within a single experiment and thus were not able to directly compare pretesting and study conditions for different delay conditions. Such comparisons, however, are crucial to understand if, and how, the size of the pretesting effect depends on retention interval.

Contrasting with the source-monitoring argument and the findings from Chan et al.'s (2018) recent meta-analysis, from a more theoretical point of view, the expectation may arise that the pretesting effect can increase with increasing retention interval. The expectation derives from the elaboration hypothesis, which is not only one of the most prominent accounts of the classic testing effect (e.g., Carpenter, 2009, 2011) but has also been suggested to explain the pretesting effect (e.g., Endres et al., 2017; Huelser & Metcalfe, 2012; see Kornell & Vaughn, 2016, for an overview of further accounts of the pretesting effect). The account assumes that attempting to retrieve a target item from memory (e.g., frog–*?*) is more likely to lead to the activation of information related to the cue item (e.g., *tongue*, *lily pad*, *fly*) than is simple study, which, on a later test, may support recall of the target item (e.g., *pond*). This hypothesis is supported, for example, by the finding that the pretesting effect typically remains absent when there is no semantic link between cue and target item (e.g., frog–*bread*; Grimaldi & Karpicke, 2012). The elaboration hypothesis also predicts that the size of the pretesting effect increases with retention interval. On the basis of the view that long-term memory is primarily semantic in nature (e.g., Bartlett, 1932) and evidence that elaborative processing can reduce forgetting (Tulving et al., 1994), the advantage of testing should be more likely to be present on a final test that is delayed rather than immediate if the testing condition enhanced elaboration (e.g., Carpenter, 2011). Initial pretesting may thus be particularly beneficial after prolonged retention interval, thus boosting the size of the pretesting effect and showing a striking parallel to the classic testing effect.

The Present Study

The aim of the two experiments reported here was to examine how prolonged retention interval modulates the pretesting effect, using both (standard) word lists consisting of weak paired associates (Experiment 1) and a more educationally relevant prose passage

(Experiment 2) as study material. In each experiment, three delay intervals between study and final retention test were applied: retention intervals of 1 min, 10 min, and 30 min in Experiment 1 and retention intervals of 1 min, 30 min, and 1 week in Experiment 2. In Experiment 1, participants studied a total of 60 weak paired associates (e.g., frog–*pond*)—half of which were pretested—and, on the final retention test, memory of all 60 associates was assessed by presenting all cue items, one by one, and asking participants to provide the correct target item (e.g., frog–*?*; e.g., Kornell et al., 2009). In Experiment 2, participants studied a prose passage containing biographical information about the former U.S. American President Woodrow Wilson. Prior to study of the passage, all subjects were asked to answer eight questions about the text in the pretesting phase and, after the retention interval, they were asked to answer 16 questions about the text on the final retention test—eight of which were already part of the pretesting phase (for a similar procedure, see Overoye et al., 2021). In both experiments, final test performance was assessed by measuring recall rates and intrusions. While we expected to find a pretesting effect for both types of study material for the 1-min retention interval—thus replicating prior work (Kornell et al., 2009; Richland et al., 2009)—it was less clear if and how the size of the pretesting effect would change across the three retention intervals. Based on the results of the meta-analysis reported in Chan et al. (2018), the magnitude of the pretesting effect may be expected to decrease with length of retention interval, whereas based on the elaboration hypothesis, the magnitude of the pretesting effect may increase with length of retention interval. The results will be of high relevance for potential application of pretesting in educational settings.

Experiment 1

Method

Ethical Considerations

All reported studies were carried out in accordance with the provisions of the World Medical Association's Declaration of Helsinki.

Participants

Using G*power (Version 3.1.9.2; Faul et al., 2009), sample size in Experiments 1 and 2 was determined on the basis of a power analysis which showed that, to detect at least a medium-sized effect for the critical interaction ($f = 0.25$), 86 subjects were to be required overall when α was set to .05 and β to .20. Closely following this recommendation, a total of 90 students ($M_{\text{age}} = 22.54$ years; 66 female, 24 male) of different German universities were recruited to participate in Experiment 1, with 30 subjects in each of the three retention interval conditions. All subjects spoke German as their native language and reported no neurological or psychiatric disease. All subjects gave their spoken informed consent and took part in the experiment in return for either course credit or a compensatory amount of money.

Material

Sixty word pairs with weak semantic associations (e.g., plate–*fork*, tradition–*Christmas*) were used as study material. The forward association strength of each word pair ranged from .051 to .053 and was .052 on average (Nelson et al., 1998). That means, when presented with the cue word, the correct target word was produced as the first response about 5% of the time. All items were translated into German. Overall, the study materials were divided into two subsets, Subset A and Subset B, consisting of 30 word pairs each. For half of the participants, Subset A was used for the study condition and Subset B for the pretest condition; for the other half of participants, the assignment of subsets to conditions was reversed.

Design

The experiment had a 2 × 3 design with the within-subjects factor of practice (study vs. pretest) and the between-participants factor of retention interval (1-min retention interval vs. 10-min retention interval vs. 30-min retention interval). Half of the pairs were presented intact and could be studied immediately (study condition), while for the other half of pairs, participants first received a test of the target word before the complete pair was shown (pretest condition). One third of the participants completed the final retention test after the 1-min retention interval, another third after the 10-min retention interval, and the remaining third after the 30-min retention interval.

Procedure

The experiment took place online and consisted of three different phases: study phase, distractor phase, and final test phase. In the study phase, half of the cue items were presented together with their target items (e.g., plate–*fork*) for 10 s and could thus be studied immediately (study condition). For the other half of the items, at first, the cue item was presented alone for 5 s, and participants were asked to guess the target item (e.g., tradition–*?*). Afterward, participants were shown the cue and target items together (e.g., tradition–*Christmas*) for another 5 s (pretest condition). Word pairs were presented in randomized order for each participant. After the study phase, all participants counted backward in steps of seven for 1 min. One third of the participants then completed the final retention test (1-min retention interval condition); another third of participants then completed the final retention test after playing the video game Tetris (https://www.geo.de/geolino/spiele/13349-rtklonlinespiel-tetris) for an additional 9 min (10-min retention interval condition); the final third of the participants completed the final retention test after playing Tetris for 9 min, engaging in the standard progressive matrices task (Raven et al., 2000) for 12 min, a spot-the-difference task for 5 min (https://www.smithsonianmag.com/games/spot-difference-180968040/), and simple arithmetic tasks for 3 min (30-min retention interval condition). On the final test, the cue words of all 60 word pairs were presented for 10 s each in a random order, and participants were asked to provide the corresponding target word (e.g., plate–?). No feedback was given during the test.[1]

[1] All study materials and data have been made publicly available at the Open Science Framework and can be accessed at https://osf.io/n7bqz/.

Results

Correct Recall

On the pretest, participants correctly guessed 3.4% of the target items. Since the effect of erroneous guesses on later studying was of main interest, those pairs in which the guess matched the correct target were excluded from further analyses.

Figure 1a shows mean recall rates on the final cued recall test as a function of practice (study vs. pretest) and retention interval (1-min retention interval vs. 10-min retention interval vs. 30-min retention interval). A 2 × 3 analysis of variance (ANOVA) with the two factors revealed main effects of practice, $F(1, 87) = 193.03$, mean square error (MSE) = 36.48, $p < .001$, $\eta_p^2 = .69$, and retention interval, $F(2, 87) = 14.33$, $MSE = 107.52$, $p < .001$, $\eta_p^2 = .25$, reflecting overall higher recall rates for pretested items than for studied items (81.4% vs. 68.9%) and an overall decrease in recall from the 1-min to the 10-min and 30-min retention interval conditions (79.7% vs. 76.2% vs. 69.7%). These main effect findings were qualified by a statistically significant interaction between the two factors, $F(2, 87) = 11.64$, $MSE = 36.48$, $p < .001$, $\eta_p^2 = .21$.

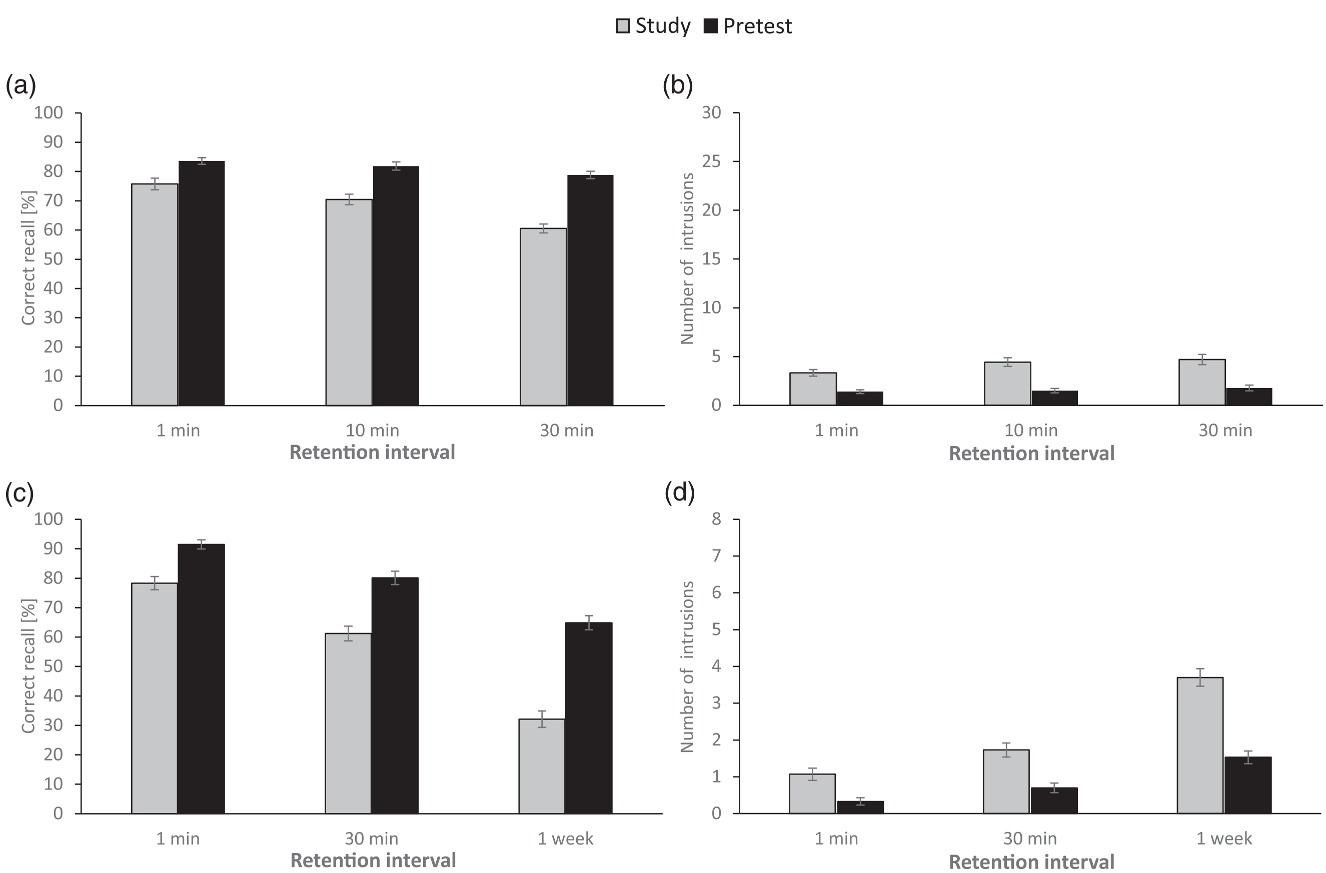
Indeed, while pairwise comparisons revealed that, relative to studied items, recall of pretested items was improved on the final test following the 1-min retention interval, $t(29) = 5.18$, $p < .001$, Cohen's $d = 0.95$; the 10-min retention interval, $t(29) = 9.26$, $p < .001$, Cohen's $d = 1.69$; and the 30-min retention interval, $t(29) = 9.77$, $p < .001$, Cohen's $d = 1.78$, the magnitude of the pretesting effect increased from 7.8% (83.6% vs. 75.8%) to 11.4% (81.9% vs. 70.5%) to 18.3% (78.8% vs. 60.6%) from the 1-min to the 10-min and 30-min retention intervals.

Intrusions

All items which participants produced during the final test that were incorrect were counted as intrusions. Figure 1b shows the number of intrusions as a function of practice (study vs. pretest) and retention interval (1-min retention interval vs. 10-min retention interval vs. 30-min retention interval). A 2 × 3 ANOVA of the two factors revealed a significant main effect of practice, $F(1, 87) = 161.57$, $MSE = 1.88$, $p < .001$, $\eta_p^2 = .65$, reflecting that, overall, the number of intrusions was higher in the study than the pretest condition (4.2 vs. 1.6 intrusions). There was no main effect of

Figure 1

Results of Experiments 1 (a, b) and 2 (c, d)



Note. (a) Recall performance on the final test of Experiment 1 (in %) as a function of practice (study vs. pretest) and retention interval (1 min vs. 10 min vs. 30 min). (b) Number of intrusions on the final test of Experiment 1 as a function of practice and retention interval. (c) Recall performance on the final test of Experiment 2 (in %) as a function of practice (study vs. pretest) and retention interval (1 min vs. 30 min vs. 1 week). (d) Number of intrusions on the final test of Experiment 2 as a function of practice and retention interval. Error bars reflect standard errors.

retention interval, $F(2, 87) = 2.03$, $MSE = 5.82$, $p = .14$, $\eta_p^2 = .05$, and no interaction between factors, $F(1, 87) = 2.66$, $MSE = 1.88$, $p = .076$, $\eta_p^2 = .06$.

Discussion

The results of Experiment 1 provide a first demonstration that the size of the pretesting effect can increase with length of retention interval. While a typical pretesting effect resulted following the short (1-min) retention interval—as reflected in improved final recall in the pretest condition relative to the study condition—the magnitude of the pretesting effect increased when the retention interval was increased to 10 min and 30 min. The results also demonstrated that, in the pretest condition, the number of intrusions on the final test is reduced relative to the study condition (e.g., Grimaldi & Karpicke, 2012) with the numerical, though not statistical, tendency that the magnitude of this benefit of pretesting increases with longer retention interval. The goal of Experiment 2 was to investigate whether the findings of Experiment 1 with weakly associated word pairs generalize to more educationally relevant study material.

Experiment 2

Method

Participants

Following Experiment 1, a total of 90 students ($M_{\text{age}} = 24.8$ years; 67 female, 23 male) of different German universities took part in Experiment 2, with 30 participants in each of the three retention interval conditions. All subjects spoke German as their native language and reported no neurological or psychiatric disease. All subjects gave their spoken informed consent and took part in the experiment in return for either course credit or a compensatory amount of money.

Material

The study material was an encyclopedically prepared German text passage about former U.S. President Woodrow Wilson. The text was 257 words long and contained both biographical facts (e.g., year and place of birth, name of parents) and political facts (e.g., name of vice president, major domestic and foreign policy achievements) about Woodrow Wilson taken from the English Wikipedia page on Woodrow Wilson (https://en.wikipedia.org/wiki/Woodrow_Wilson). The test questions were constructed based on 16 facts identified as testable items in the text passage. These 16 questions were divided into two subgroups (Subgroup A and Subgroup B). 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?"). For half of the participants, the pretest contained the questions of Subgroup A, and for the other half, the pretest contained the questions of Subgroup B. All 16 questions were included in the final test and presented in a random order. Thus, eight questions on the final test previously had been subject to a pretest (pretest items) and eight questions had not been previously tested (study items).

Design

The experiment had a 2 × 3 design with the within-subjects factor of practice (study vs. pretest) and the between-subjects factor of retention interval (1-min retention interval vs. 30-min retention interval vs. 1-week retention interval). Conditions differed in whether the target items were only presented in the text passage (study condition) or had been asked in a pretest before study (pretest condition). Retention interval differed in whether the final test took place 1 min, 30 min, or 1 week after the study phase. Thirty participants each were randomly assigned to each of the three retention interval conditions.

Procedure

In the study phase, participants were told that they would be presented a text about Woodrow Wilson to study for a later memory test. Before the text was shown, participants were asked eight open-answer questions about Woodrow Wilson (e.g., "In which state was Woodrow Wilson born?") and were instructed to provide an answer to all questions, regardless of whether they knew the answer. Each of those questions lasted on the screen until an answer was given. After the pretest was finished, the participants were presented the text passage for 5 min. Next, all participants were instructed to count backward aloud in steps of seven for 1 min, after which one third of participants completed the final test (1-min retention interval condition); a second third of participants continued working on the same additional distractor tasks as in the 30-min retention interval condition of Experiment 1 before they engaged in the final test (30-min retention interval condition); the final third of participants was finished for the time being after the backward counting task, but were asked to complete the final test on a second appointment 1 week later (1-week retention interval condition).[2] On the final test, all 16 questions (8 pretested questions and 8 new questions) were presented in a random order. Every question remained on the screen until an answer was given.

Results

Recall Performance

On the initial pretest that preceded the presentation of the text passage, participants correctly guessed 2.8% of the questions. Like in Experiment 1, all items that were answered correctly on the pretest were removed from further analyses.

The percentage of correctly answered questions on the final test is shown in Figure 1c as a function of practice (study vs. pretest) and retention interval (1-min retention interval vs. 30-min retention interval vs. 1-week retention interval). A 2 × 3 ANOVA of the two factors revealed main effects of practice, $F(1, 87) = 347.04$, $MSE = 60.54$, $p < .001$, $\eta_p^2 = .80$, and retention interval, $F(2, 87) = 76.97$, $MSE = 262.75$, $p < .001$, $\eta_p^2 = .64$, reflecting higher overall recall rates for pretested items than for studied items (78.8% vs. 57.2%) and higher recall rates when the final test took place after 1 min than after 30 min or 1 week (84.9% vs. 70.7% vs. 48.5%). There was also a significant interaction between the two factors, $F(2, 87) = 25.29$, $MSE = 60.54$, $p < .001$, $\eta_p^2 = .37$.

[2] All subjects who attended the first session of the 1-week retention interval condition returned for the second session.

Indeed, while pairwise comparisons revealed that, relative to studied items, recall of pretested items was improved on the final test following the 1-min retention interval, $t(29) = 6.94$, $p < .001$, Cohen's $d = 1.27$; the 30-min retention interval, $t(29) = 10.07$, $p < .001$, Cohen's $d = 1.84$; and the 1-week retention interval, $t(29) = 14.66$, $p < .001$, Cohen's $d = 2.68$, the magnitude of the pretesting effect increased from 13.2% (91.5% vs. 78.3%) to 18.9% (80.1% vs. 61.3%) to 32.8% (64.9% vs. 32.1%) from the 1-min to the 30-min and 1-week retention intervals.

Intrusions

The mean number of intrusions on the final test is shown in Figure 1d as a function of practice (study vs. pretest) and retention interval (1-min retention interval vs. 30-min retention interval vs. 1-week retention interval). A 2 × 3 ANOVA of the two factors revealed main effects of practice, $F(1, 87) = 162.17$, $MSE = 0.48$, $p < .001$, $\eta_p^2 = .65$, and retention interval, $F(2, 87) = 45.44$, $MSE = 1.30$, $p < .001$, $p < .001$, $\eta_p^2 = .51$, reflecting that, overall, there was a higher number of intrusions for studied items than for pretested items (2.2 vs. 0.9) and fewer intrusions were made after 1 min than 30 min and 1 week (0.7 vs. 1.2 vs. 2.6). There was also a significant interaction between the two factors, $F(2, 87) = 17.97$, $MSE = 0.48$, $p < .001$, $\eta_p^2 = .29$. In fact, while pairwise comparisons revealed that, relative to studied items, fewer intrusions were made for pretested items following the 1-min retention interval, $t(29) = 4.85$, $p < .001$, Cohen's $d = 0.89$, the 30-min retention interval, $t(29) = 6.10$, $p < .001$, Cohen's $d = 1.11$, and the 1-week retention interval, $t(29) = 10.35$, $p < .001$, Cohen's $d = 1.89$, the magnitude of this benefit of pretesting increased from 0.7 (1.1 vs. 0.3) to 1.0 (1.7 vs. 0.7) to 2.2 (3.7 vs. 1.5) from the 1-min to the 30-min and 1-week retention intervals.[3]

Discussion

The findings of Experiment 2 extend the results of Experiment 1 by showing that the delay-induced increase in the size of the pretesting effect is not restricted to the case when weak associates are applied as study material, but can also arise for more complex prose passages, even when length of retention interval is extended to 1 week. Indeed, while a reliable pretesting effect already arose after the 1-min retention interval, there was an increase in the size of the pretesting effect following the 30-min retention interval and, in particular, the 1-week retention interval—for which a massive Cohen's d effect size of 2.68 was observed. The results of Experiment 2 also showed that in the pretest condition, the number of intrusions on the final test was reduced relative to the study condition. Again, this benefit for the pretested information was affected by retention interval and increased reliably after the 30-min and 1-week retention intervals. This finding replicates a pattern that was also present in Experiment 1, though only numerically and not statistically.

General Discussion

Employing both weak associates (Experiment 1) and a prose passage (Experiment 2) as study material, the present study found that the size of the pretesting effect can increase with retention interval. Strikingly, the size of the effect roughly doubled when the retention interval increased from 1 min to 30 min in Experiment 1, and when the retention interval increased from 1 min to 1 week in Experiment 2. The present study thus provides the first demonstration that the beneficial effects of pretesting can become even more pronounced when information needs to be retained over longer periods of time. The results of both experiments also indicated that, relative to additional study, pretesting can reduce the number of intrusions produced on the final test, and this intrusion-reducing effect of pretesting become even more distinct with prolonged retention interval.

At first glance, the present findings may appear inconsistent with the results of the meta-analysis reported by Chan et al. (2018), which suggested that the size of the pretesting effect may slightly decrease with retention interval. However, this meta-analysis did not include studies in which length of retention interval was manipulated within single experiments. Moreover, in the data sets that were included, length of retention interval did not exceed 15 min, with the only exception of Kornell et al. (2009, Experiment 5) in which delay between study and test was 38 hr. Interestingly, visual inspection of Figure 6d in Chan et al. (2018) suggests that, in Kornell et al.'s (2009) Experiment 5, the size of the pretesting effect that occurred after the 38-hr delay was substantially larger than the mean effect size arising for the 0–15 min delays in this meta-analysis, which is consistent with the results reported here. The fact that the meta-analysis found no overall increase in the pretesting effect with retention interval thus likely is caused by a lack of data sets for longer retention intervals.

The results of the present study also do not seem fully consistent with the only other study in the literature that applied a within-experiment manipulation of retention interval (Overoye et al., 2021). Like in the present Experiment 2, subjects in Overoye et al.'s Experiment 2 studied a prose passage after they were asked to answer eight prequestions on the text. A final retention test showed a typical pretesting effect with enhanced recall of pretested, relative to nonpretested, information. While the pretesting effect arose both for subjects for whom the test was conducted after a 5-min retention interval and a 1-week retention interval, the magnitude of the effect did not increase with delay, unlike in our study. Critically, however, the prose passages used in this study were almost three times as long (approximately 750 words) as in the present Experiment 2, which led to much lower mean recall rates for the pretested and nonpretested items after the 1-week retention interval (23% vs. 16%) than in the present Experiment 2 (65% and 32%). Conceivably, the low recall rate of 16% in the study condition reflects a floor effect and, if so, may have prevented the recall difference between pretesting and study to increase from short to prolonged retention interval, which may explain why the pretesting effect increased with delay in the present study, but not in this

[3] For both experiments, we also examined for the pretesting condition how often participants' initial guesses during pretesting came up as intrusions on the final test. In Experiment 1, in which subjects were pretested on 30 cue–target pairs, mean number of "guess" intrusions was 0.40, 0.37, and 0.50 for the 1-min, 10-min, and 30-min retention interval conditions. Averaged across delay conditions, guess intrusions thus made up ca. 25% of all intrusions. In Experiment 2, in which subjects were pretested on eight facts from the prose passage, mean number of guess intrusions was 0.00, 0.07, and 0.23 for the 1-min, 30-min, and 1-week retention interval condition. Averaged across delay conditions, guess intrusions thus made up ca. 10% of all intrusions. The relatively rare occurrence of guess intrusions seems to suggest that participants are quite good at distinguishing their initial guesses from the target items to be retrieved.

prior study. Future work may address the issue and examine these conditions in more detail in the absence of any floor effects.

The present findings suggest a parallel between the pretesting effect and the classic testing effect, both of which apparently increase with retention interval. This parallel is consistent with the view that both testing effects may be mediated by elaboration processes (Carpenter, 2009, 2011; Huelser & Metcalfe, 2012). For the classic testing effect, the proposal is that when during the initial testing phase, subjects are shown a cue item and attempt to retrieve the previously studied target item, memory representations associated to the cue item are activated and become related to the cue–target pairing, given that the target item has been successfully retrieved. At the final retention test, this elaborated memory trace provides further retrieval paths through which the target item can be accessed. For the pretesting effect, it has been suggested that when, during initial pretesting, subjects are presented with a cue item and attempt to guess the target item, memory representations related to the cue–target pairing are activated and thus create a more elaborate memory trace through which, at the later test, access to the target item is facilitated. On the basis of the assumption that recall may become more semantic with increasing delay (Carpenter, 2011) for both the classic testing effect and the pretesting effect, the elaboration hypothesis can explain the observed increase in the size of the testing effect.

While the present results are consistent with an elaboration account of the pretesting effect, like results on the classic testing effect, they are not necessarily inconsistent with an alternative contextual account (Karpicke et al., 2014; Metcalfe & Huelser, 2020; Overman et al., 2021). When encoding information, we also store features of the temporal context that is present when we encounter that information (Estes, 1955; Mensink & Raaijmakers, 1988). Pretested items could therefore get associated with an enriched temporal context—encompassing an integrated representation of the pretesting context and the study context—whereas studied items may get associated with the study context only (Hunt et al., 2011; Pierce et al., 2017). Pretested material may benefit from the enriched temporal context at test, particularly after longer delay when the test context has become increasingly dissimilar to the study context and recall benefits from strong contextual cues. Consequently, pretested items may have a growing recall advantage over studied items with delay, as was observed in the present study. Critically, as is true for the classic testing effect, elaboration and contextual accounts are not mutually exclusive and both types of mechanisms may have contributed to the observed increase of the effect with prolonged retention interval.

Like posttesting in the classic testing effect, pretesting provides a promising method to retain study information over an extended period of time in educational settings. Although some educators may be hesitant to incorporate pretesting as a learning tool into their everyday teaching because they initially lead to numerous wrong answers, the present results may alleviate such concerns: Generating errors early can increase the number of correct answers and reduce incorrect answers in the long run. All this holds while educators should still consider that the memorial benefits of pretesting may be limited to the pretested information, as typically no reliable beneficial effects of pretesting on other information within the same text passage have been found (Carpenter et al., 2018; Hausman & Rhodes, 2018; James & Storm, 2019; Richland et al., 2009; but see Carpenter & Toftness, 2017; Little & Bjork, 2016). A related issue is that of transfer. In the current Experiment 2, the same test format was used on the initial pretest and the final test, and the questions used during pretesting were also worded exactly like the questions on the final test. While prior research on vocabulary learning has already shown that pretesting can still be beneficial when the initial and final test formats differ (Potts et al., 2019; Potts & Shanks, 2014), thus suggesting at least some level of transfer for pretested information, future research is still required to examine whether different wording of questions during initial and final testing also yields a reliable pretesting effect.

Conclusions

The beneficial effects of pretesting—as reflected in an increase in correct recall of studied information and a reduction in number of intrusions—become even more pronounced when the retention interval between study and final testing of the information is prolonged. This pattern holds both when weak associates are used as study material and the retention interval lasts up to 30 min, and when a prose passage is used as study material and the retention interval lasts up to 1 week. The findings suggest that pretesting may be an effective tool to increase learning, especially in settings featuring longer periods of time between acquisition and the final testing of study materials.

References

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review*, *33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

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

Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition*, *6*(1), 104–109. https://doi.org/10.1016/j.jarmac.2016.07.014

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

Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, *49*, 13–20. https://doi.org/10.1016/j.learninstruc.2016.11.010

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

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

Hausman, H., & Rhodes, M. G. (2018). When pretesting fails to enhance learning concepts from reading texts. *Journal of Experimental Psychology: Applied*, *24*(3), 331–346. https://doi.org/10.1037/xap0000160

Hilaire, K. J. S., & Carpenter, S. K. (2020). Prequestions enhance learning, but only when they are remembered. *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*(4), 514–527. https://doi.org/10.3758/s13421-011-0167-z

Hunt, R. R., Smith, R. E., & Dunlap, K. R. (2011). How does distinctive processing reduce false recall? *Journal of Memory and Language*, *65*(4), 378–389. https://doi.org/10.1016/j.jml.2011.06.003

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

Johnson, M. K. (1997). Source monitoring and memory distortion. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *352*(1362), 1733–1745. https://doi.org/10.1098/rstb.1997.0156

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

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

Karpicke, J. D., Lehman, M., & Aue, R. W. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 61, pp. 237–284). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 106–114. https://doi.org/10.1037/a0033699

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

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition*, *44*(7), 1085–1101. https://doi.org/10.3758/s13421-016-0621-z

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*, 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*. http://w3.usf.edu/FreeAssociation/

Overman, A. A., Stephens, J. D. W., & Bernhardt, M. F. (2021). Enhanced memory for context associated with corrective feedback: Evidence for episodic processes in errorful learning. *Memory*, *29*(8), 1017–1042. https://doi.org/10.1080/09658211.2021.1957937

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

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning*. National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. https://doi.org/10.1037/e607972011-001

Pierce, B. H., Gallo, D. A., & McCain, J. L. (2017). Reduced interference from memory testing: A postretrieval monitoring account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1063–1072. https://doi.org/10.1037/xlm0000377

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. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–667. https://doi.org/10.1037/a0033194

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

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, & 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, 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

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/10.1016/j.jarmac.2018.06.003

Tulving, E., Kapur, S., Craik, F. I., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. *Proceedings of the National Academy of Sciences*, *91*(6), 2016–2020. https://doi.org/10.1073/pnas.91.6.2016

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 July 8, 2022
Revision received September 23, 2022
Accepted October 19, 2022 ■