Psychology and Aging

Making Guesses During Learning Can Be Beneficial for Older Adults' Memory

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

Online First Publication, July 31, 2025. https://dx.doi.org/10.1037/pag0000929

CITATION

Kliegl, O., Bartl, J., & Bäuml, K.-H. T. (2025). Making guesses during learning can be beneficial for older adults' memory. *Psychology and Aging*. Advance online publication. https://dx.doi.org/10.1037/pag0000929

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology and Aging

ISSN: 0882-7974

https://doi.org/10.1037/pag0000929

Making Guesses During Learning Can Be Beneficial for Older Adults' Memory

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

When young adults complete a pretest (e.g., star—?) before to-be-learned material is studied (e.g., star—*night*), their recall of the material is typically enhanced on a subsequent final test compared to material studied without initial pretesting. The present two experiments investigated whether this pretesting effect arises also in older adults and is modulated in size when repeated guessing attempts are made during pretesting. Sixty young adults (mean age = 24.5 years) and 60 older adults (mean age = 68.6 years) took part in Experiment 1, while 54 young adults (mean age = 21.8 years) and 54 older adults (mean age = 66.6 years) took part in Experiment 2. Results showed that, like young adults, older adults can benefit from a single guessing attempt made during pretesting, both when weakly associated word pairs (Experiment 1) and prose passages (Experiment 2) were used as study material. However, multiple guessing attempts during pretesting led to an additional recall benefit when word pairs but not when prose passages had been studied. Experiment 2 also examined possible transfer effects of pretesting and showed a lack of transfer to previously studied but untested information, for both young and older adults. The results are discussed with respect to prominent accounts of the pretesting effect.

Public Significance Statement
Previous research has shown that taking a test on information (e.g., on a topic such as the former U.S. President Woodrow Wilson) even before studying the information can improve later recall on an assessment test. We show that such pretesting can increase correct recall and decrease incorrect recall for both young and older adults and, under some conditions, can lead to further improvements when multiple unique guesses are made during the pretest. The findings suggest that pretesting may be an effective learning technique not only for young adults but also for older adults.

Keywords: testing effect, pretesting effect, young and older adults, elaboration account, cue-overload principle

Testing during learning can promote memory. Over the last 2 decades, research on the testing effect has impressively shown that memory tests not only serve to determine the current level of knowledge but also represent learning events themselves. Retrieval practice of previously encoded information indeed often causes better retention of this information in a subsequent final test than does restudy of the information, with a wide variety of both laboratory and classroom studies demonstrating that the testing effect is a robust memory phenomenon (for reviews, see Karpicke, 2017; Rowland, 2014).

Strikingly, prior work on the pretesting effect has shown that testing even *before* the material to be learned is exposed for study can benefit later retention of the material for young adults (for reviews, see Kornell & Vaughn, 2016; Pan & Carpenter, 2023). In a typical pretesting effect task, participants engage in a series of study-only trials in which they are asked to learn weakly associated cue-target pairs (e.g., frog–*pond*) which are randomly intermixed with a series of pretest trials in which, at first, only the cue item of a pair is presented for a few seconds and the target item has to be guessed (plate–*?*) before it is revealed (plate–*fork*). On a later final test, in which participants are presented, one by one, with all prior cue items and asked to recall the associated target item (plate–*?*), memory performance is usually better for target items that were originally pretested compared to target items that were studied only. The pretesting effect even arises if only those trials in which participants made an incorrect guess during pretesting are included

Hannes Zacher served as action editor.

Oliver Kliegl https://orcid.org/0000-0001-7730-4019

The ideas and data appearing in the article have not been disseminated before. All study materials and data have been made publicly available on the Open Science Framework and can be found at https://osf.io/uge3r/. This work was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft) awarded to Oliver Kliegl (Grant KL 2967/2-1).

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

Correspondence concerning this article should be addressed to Oliver Kliegl, Department of Experimental Psychology, Regensburg University, Universitaetsstrasse 31, 93040 Regensburg, Germany. Email: oliver.kliegl@ur.de

in further analyses (e.g., Kornell et al., 2009), which suggests that learning from errors is indeed possible. Research over the past 15 years has demonstrated the robustness of the pretesting effect for a variety of study materials—including weakly associated word pairs, videos, trivia questions, and foreign language vocabulary—in both laboratory and classroom settings (for reviews, see Chan et al., 2018; St. Hilaire et al., 2024).

One prominent explanation of the pretesting effect is the elaboration hypothesis (e.g., Huelser & Metcalfe, 2012). This account proposes that when subjects are shown a cue item (plate–*?*) in an initial pretest and are asked to guess the target item, memory representations that are semantically related to the cue item—and may include items such as *table*, *spoon*, *knife*—become activated and linked to the pair when correct-answer feedback is given a few seconds later. When, on the final test, the cue item is provided and the target item has to be produced, these additional memory representations—among which is also the incorrect guess from the pretest—may act as semantic mediators which facilitate recall of the target item, thereby enabling the pretesting effect. Huelser and Metcalfe provided support for the elaboration account by showing that pretests not only enhance later memory of the target item but also increase response times for the target items. In particular, attempting to retrieve an originally pretested item may initially result in the retrieval of mediating memory representations, thereby slowing down access to the target item. Further evidence in favor of the account comes from a recent study showing that the benefits of pretesting on later recall performance were greater when three initial guessing attempts were made than when only one guessing attempt was made (Kliegl et al., 2024). Multiple unique guesses may induce more elaborations than only a single guess and could act as additional mediating information through which retrieval of the target information is enabled on the later test, thus leading to the observed increase in the size of the pretesting effect.

At first glance, it may be surprising that making mistakes during learning can boost memory. In fact, a large body of classic research on errorless learning has demonstrated that allowing for errors during the learning process can hinder the retention of information (e.g., Baddeley & Wilson, 1994; Skinner, 1957). For decades, educators therefore have been advocating for the use of errorless learning as a critical educational tool, by emphasizing the importance of creating study conditions which eliminate or minimize errors in any learning contexts (e.g., Duker, 1981; Keel & Gast, 1992). Furthermore, clinicians have recommended errorless learning as an effective learning method for those with memory impairments, including older adults, or individuals with amnesia or dementia (for a review, see Clare & Jones, 2008). The errorless learning perspective also aligns well with a variety of memory models which include the cue-overload principle (e.g., Raaijmakers & Shiffrin, 1981; Watkins & Watkins, 1975). The principle asserts that the probability of retrieving information decreases the more information gets linked to an associated cue item. Since the added information can also involve erroneous guessing attempts produced on the pretest, the guesses may interfere with the retrieval of target information on the final test, therefore decreasing memory of the target item.

A number of studies indeed have found that pretesting older adults can impair their later recall performance (e.g., Wilson et al., 1994), which, at first glance, seems to suggest that pretesting may be generally beneficial only for younger adults. However, these earlier studies have typically used study material for which the subject-generated error was semantically unrelated to the correct response, for example, participants may be presented with a word stem such as "qu___" and guess the word "queen" before they learn that the correct response is "quote." Consistently, prior studies have found also for young adults that the pretesting effect can be reduced or even reversed when the generated errors are unrelated to the correct response (e.g., Zawadzka & Hanczakowski, 2019). This finding is in line with the elaboration account since under such circumstances it should be difficult for participants to properly integrate the guess with the cue-target pair, so that the guess may hinder rather than help access to the target item on the final test.

Consequently, the suggestion arising from prior work that young, but not older, adults can benefit from pretesting may be due to the fact that studies with young adults have tended to use study materials that allow for better integration of the generated errors with the target information. In support of this possibility, Cyr and Anderson (2012, 2015) showed that, in both young and older adults, pretesting was able to boost subsequent correct recall and source memory when during initial acquisition, conceptual cues (i.e., category names such as "fruit") were presented and participants were asked to guess the associated target item. These researchers also showed that when a lexical cue was presented during acquisition, that is, a word stem such as "fr___" and the word had to be guessed, neither young nor older adults benefited from such pretesting (Cyr & Anderson, 2015). Since lexical cues usually lead to guesses that are unrelated to the target items (e.g., a participant may guess the word *friend* with the correct answer being *frozen*), the conclusion appears to be that older adults can benefit from pretesting as much as young adults as long as conceptual cues are used during pretesting.

Such conclusion might be premature, however. The one reason is that Cyr and Anderson (2012, 2015) did not isolate the effects of initial erroneous guesses on later recall performance. Rather, final test performance in the two studies was based on trials for which the target items were *and* were not guessed correctly during pretesting. The two studies also did not report how often correct guesses were made on the initial pretest trials and how common the target items were for a given category. Including pairs for which correct guesses were made initially makes it impossible to isolate the effects of initially generated errors on later recall performance (for a more detailed discussion of the issue, see Kornell et al., 2009). Therefore, one of the major aims of the present study was to investigate whether older adults can still benefit from pretesting when all trials on which correct guesses were made during pretesting are not included in the analysis of final test performance. Furthermore, the present two experiments employed study material for which the chance of correct guessing is relatively low (i.e., ≤5%).

The second reason is that the two Cyr and Anderson (2012, 2015) studies always had participants make two guesses on each pretest trial. It, therefore, remains to be investigated whether a single guessing attempt can be sufficient to produce the pretesting effect in older adults, and, in particular, whether the magnitude of the pretesting effect in older adults increases with the number of initial guesses, as it does in younger adults. Since older adults have been shown to be more vulnerable to memory interference than young adults (e.g., Kliegl et al., 2019), it is possible that multiple guesses made during pretesting might impede, rather than facilitate, retrieval of the target information on the final test. The present study thus examined whether initial guessing can lead to parallel benefits in young and older adults that go beyond demonstrating the mere existence of a pretesting effect.

The Present Study

Two experiments were conducted to compare the effects of pretesting on later recall performance between young and older adults. While Experiment 1 used weakly related word pairs (frog—*pond*) as study material, Experiment 2 used more educationally relevant prose passages. The first goal of the two experiments was to examine whether older participants can benefit from pretesting in the typical pretest situation in which only one initial guessing attempt is made. The second goal was to replicate prior findings for young adults that the magnitude of the pretesting effect can increase when multiple initial guesses are made and to determine whether this benefit of repeated guessing attempts generalizes to older adults. Since older adults are particularly vulnerable to interference from competing items (e.g., Kliegl et al., 2019), multiple additional guessing attempts indeed may impede, rather than promote, access to the target information on the final test in this age group. For instance, older adults may be able to use one guessing attempt as a stepping stone to access the target information on the later test, but two additional guessing attempts may not be properly integrated with the cue-target pair and rather serve as a source of interference when the target information needs to be retrieved. The present study examined this possibility. The third goal was to determine whether initial guessing can induce memory transfer to previously studied material that is not pretested. This issue will be examined in Experiment 2 which uses prose passages as study material and in which, during acquisition of a passage, the presentation of untested material occurs in close temporal proximity to the pretests. The assumption is that potentially short-lived elaboration processes might spill over to the untested material and thus boost later retention of the untested material.

Experiment 1

Experiment 1 investigated for both young and older adults the effects of a single initial guessing attempt and of three initial guessing attempts on subsequent recall performance when using weakly associated word pairs as study material. Therefore, the two age groups were either presented with both the cue item and target item of a pair from the beginning of the trial (study-only condition), or they were only provided with the cue item and instructed to make one guess (one-guess condition) or three guesses (three-guesses condition) before the target item was shown. The study material consisted of weakly linked cue-target pairs (e.g., frog—*pond*). In the study-only condition, the complete word pair was presented intact for 10 s (e.g., frog—*pond*). In the one-guess condition, only the cue item was provided for the first 5 s of a trial (e.g., frog—*?*) and a single guess had to be made within that time span before the intact word pair was shown for another 5 s. The three-guesses condition was similar to the one-guess condition, with the critical difference that, after providing their first guess, a second and a third guess had to be made. Participants were allotted 5 s each for these additional guesses. After this acquisition phase, a final test was conducted on which memory of all target items was assessed (e.g., frog—*?*). Both correct recall and commission errors (i.e., intrusions) were measured on this final test.

Method

Transparency and Openness

The deidentified data on which the study conclusions are based and the materials are available. The link to access this information is provided in the author note. The analytic code is not available since all statistical analyses reported in the article can be reproduced on the basis of the details given in the Method sections of Experiments 1 and 2. We report how we determined our sample size, any data exclusions, all manipulations, and all measures. None of the reported experiments have been preregistered.

Participants

Optimal sample sizes for Experiments 1 and 2 were assessed with a power analysis using the G*Power software (Version 3.1.9.2; Faul et al., 2009). On the basis of an estimate of the meta-analytic effect size of the pretesting effect (Hedge's $g = 0.44$, Boustani & Shanks, 2022), the analysis revealed that 43 subjects were required overall when setting α to .05 and β to .20. To detect a potentially smaller benefit of pretesting in older adults, we recruited a higher number of participants for each age group: The final sample consisted of 60 young adults with a mean age of 24.5 years (44 female, 16 male, zero diverse) and 60 older adults with a mean age of 68.6 years (41 female, 19 male, zero diverse). We recruited young adults from Regensburg University and older adults from the community. All data were collected between April 2023 and August 2023. All participants were Caucasian and gave their spoken informed consent. Young participants were compensated either financially or earned some course credit for their participation. Older participants all received a compensatory amount of money for their participation. Compared to young adults, older adults performed significantly worse in the Mini Mental State Examination (Folstein et al., 1975; young adults: 29.5; older adults: 28.1), $t(118) = 7.512$, $p < .001$, Cohen's $d = 1.37$, and the digit span task (young adults: 15.4; older adults: 13.3), $t(118) = 4.32$, $p < .001$, Cohen's $d = 0.79$, but significantly better in the MWT-B (Mehrfachwahl Wortschatz Test [multiple-choice vocabulary intelligence test]; Lehrl, 2005), a German vocabulary test that assesses crystallized intelligence (young adults: 27.7; older adults: 31.0), $t(118) = 6.09$, $p < .001$, Cohen's $d = 1.11$.

Material

Sixty weakly related word pairs (e.g., plate—*fork*, frog—*pond*) were employed as study material. We used cue-target pairs from the Nelson et al.'s (1998) norms and selected them on the basis of a relatively narrow forward association strength of .051 to .053. The average association strength was .052, meaning that participants in the Nelson et al.'s study came up with a target item as their first associate approximately 5% of the time when a given cue item was shown. All 60 pairs were translated into German and divided into three subsets of 20 pairs each. Each subset was applied equally often in each of the experimental conditions.[1]

Design

The experiment used a mixed design with the between-subjects factor AGE GROUP (young adults vs. older adults) and the within-subjects factor TYPE OF PRACTICE (study only vs. one guess vs. three guesses). TYPE OF PRACTICE was manipulated by either presenting the cue-target pairs intact from the beginning of the trial (study-only

[1] The study materials and data can be found at https://osf.io/uge3r/.

condition), or by providing only the cue item and having participants come up with one guess (one-guess condition) or three unique guesses (three-guesses condition) before revealing the target item.

Procedure

The protocol employed in this study was deemed exempt by the ethical review board of Regensburg university. The experiments were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. For both young and older participants, the experiment consisted of three phases: an acquisition phase, a distractor phase, and a final test phase. While all 60 word pairs were presented during the initial acquisition phase, only 20 of these pairs were shown intact from the start of a trial. In these study-only trials, the cue-target pair was presented for 10 s (plate—*fork*) and could therefore be encoded immediately. All of the remaining 40 trials involved some form of pretesting. Twenty of the pairs were allocated to one-guess trials in which, within a time frame of 5 s, the target item had to be guessed from the cue item (e.g., tradition—*?*). Immediately prior to the start of each trial, a tone was presented for 0.5 s to indicate that a guessing attempt had to be produced. The remaining 20 pairs were allocated to three-guesses trials in which subjects were given two further 5 s periods to make the additional guesses. Presentation of the tone occurred immediately prior to the start of each 5-s period to indicate that another guessing attempt was required. Subjects were not informed at the beginning of each trial how many guesses they would have to make. At the end of the one-guess and three-guesses trials, the complete cue-target pairs were presented for 5 s (e.g., tradition—*Christmas*). Presentation order of the word pairs was random for all participants. There was no consequence if participants were unable to make a guess within the 5 s time frame. Indeed, prior work has shown that final test recall is largely unaffected by whether people are forced to guess in the pretest or are allowed to leave answers blank (Kang et al., 2011; Metcalfe & Kornell, 2007).

After the acquisition phase, subjects engaged in the standard progressive matrices task (Raven et al., 2000) as a distractor activity for 5 min before the final test was conducted. In this test, the cue items of all 60 word pairs were shown in randomized order for 10 s each. At the beginning of each 10 s period, the 0.5 s tone was presented to indicate to participants that they should provide the corresponding target item. Participants did not receive any feedback on this final test.

Results

Initial Pretest

For both Experiments 1 and 2, we applied the data aligned rank transform (ART) procedure of the ARTool package (Kay et al., 2021; Wobbrock et al., 2011) using R (Version 2024.12.1) to conduct nonparametric analyses of variance (ANOVAs) on the pretest and final test data.[2] A 2 × 2 ANOVA of the between-subjects factor AGE GROUP (young vs. older adults) and the within-subjects factor TYPE OF PRACTICE (one guess vs. three guesses) on percentage of correct guesses during the initial pretest showed significant main effects of AGE GROUP, $F(1, 118) = 21.53$, $p < .001$, $\eta_p^2 = .15$, and TYPE OF PRACTICE, $F(1, 118) = 57.55$, $p < .001$, $\eta_p^2 = .33$, reflecting that, overall, young adults made more correct guesses than older adults (7.0% vs. 4.0%) and more correct guesses were made for item pairs for which participants were allowed to make three guessing attempts, relative to one guessing attempt (7.9% vs. 3.1%). There was also a significant interaction between factors, $F(1, 118) = 10.15$, $p = .001$, $\eta_p^2 = .08$. While planned comparisons using the Wilcoxon signed-rank test showed that, relative to the one-guess condition, correct guesses were increased in the three-guesses condition for both age groups, all ps ≤ .002, this difference was more pronounced in young than older adults (6.8% vs. 2.8%).

Final Test

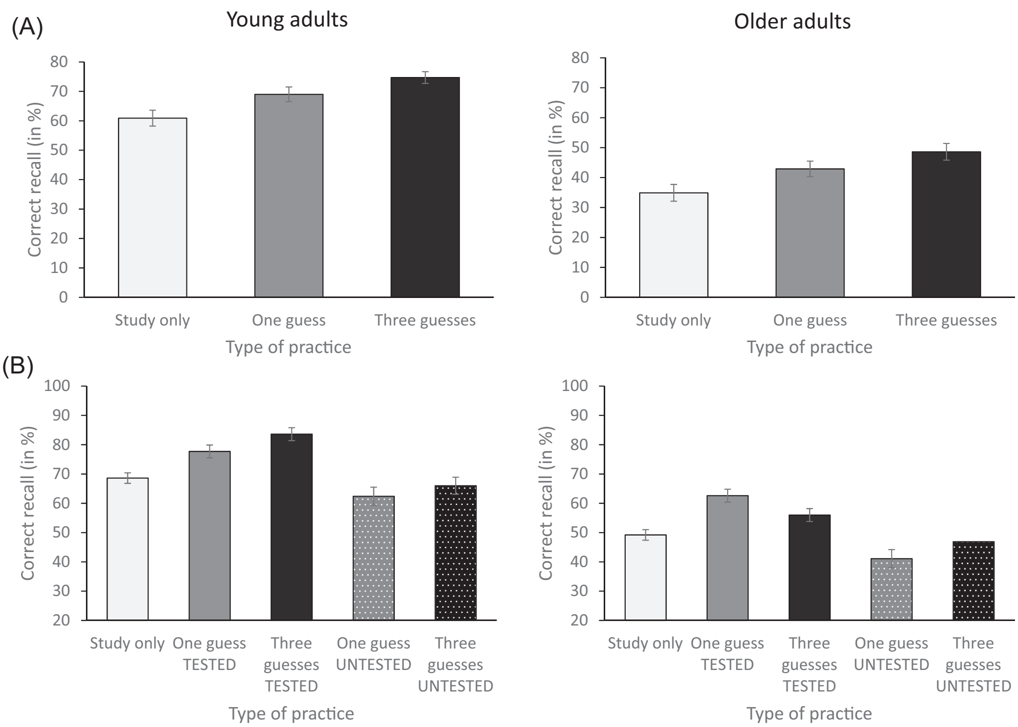
Correct Recall. Figure 1A shows the percentage of correctly recalled items on the final test as a function of the between-subjects factor AGE GROUP (young adults vs. older adults) and the within-subjects factor TYPE OF PRACTICE (study only vs. one guess vs. three guesses). A 2 × 3 ANOVA of the between-subjects factor AGE GROUP (young vs. older adults) and the within-subjects factor TYPE OF PRACTICE (study only vs. one guess vs. three guesses) on percentage of correct responses on the final test revealed significant main effects of AGE GROUP, $F(1, 118) = 70.84$, $p < .001$, $\eta_p^2 = .38$, and TYPE OF PRACTICE, $F(2, 236) = 33.09$, $p < .001$, $\eta_p^2 = .22$, reflecting that, overall, recall of older adults was lower than of young adults (42.1% vs. 68.2%) and that, overall, recall performance increased with number of initial guessing attempts (47.9% vs. 55.9% vs. 61.7%). Critically, there was no significant interaction between the two factors, $F(2, 236) < 1$, $\eta_p^2 = .0006$, suggesting that the effects of TYPE OF PRACTICE did not differ reliably for the two age groups.

Two separate 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on correct-recall performance were conducted, with the first ANOVA including only the study-only and one-guess conditions, and the second ANOVA including only the one-guess and three-guesses conditions. Critically, both ANOVAs again revealed significant main effects of TYPE OF PRACTICE, all ps < .001, reflecting that overall recall performance was improved in the one-guess condition, relative to the study-only condition, and was improved in the three guesses-condition, relative to the one-guess condition. Both ANOVAs showed no interaction between factors, all ps ≥ .85.

Intrusions. All commission errors produced on the final test were counted as intrusions. Table 1 shows intrusion rates—that is, the number of intrusions divided by the number of study items in a condition—as a function of AGE GROUP and TYPE OF PRACTICE. A 2 × 3 mixed ANOVA of the two factors on intrusion rate revealed significant main effects of AGE GROUP, $F(1, 118) = 6.58$, $p = .012$, $\eta_p^2 = .05$, and TYPE OF PRACTICE, $F(2, 236) = 15.62$, $p < .001$, $\eta_p^2 = .12$, reflecting that the overall intrusion rate of older adults was higher than of young adults (30.4% vs. 21.2%) and that the overall intrusion rate decreased with number of initial guessing attempts (30.1% vs. 25.0% vs. 22.2%). Critically, there was no significant interaction

[2] Nonparametric ANOVAs were used because (a) the high frequency of zero correct guesses on the pretest can pose an issue for parametric ANOVA with its normality assumption and (b) the number of items on which the calculation of final test recall is based on can vary slightly between participants, which a parametric ANOVA cannot account for. Both issues are unproblematic with the nonparametric ART procedure (Wobbrock et al., 2011). We thank an anonymous reviewer for pointing out the potential problems when using parametric ANOVAs for this type of data analysis but would also like to emphasize that for both experiments exactly the same pattern of results emerged when parametric ANOVAs were used.

Figure 1
Results of Experiments 1 and 2



Note. (A) Results of Experiment 1. Recall performance on the final test (in %) for young and older participants for each of the three types of practice (study only vs. one guess vs. three guesses). (B) Results of Experiment 2. Recall performance on the final test (in %) for young and older participants for the study-only condition and, separately for tested and untested sentences, for the one-guess and three-guesses conditions. Error bars reflect standard errors.

between the two factors, $F(2, 236) = 1.21$, $p = .30$, $\eta_p^2 = .01$, suggesting that the effects of TYPE OF PRACTICE did not differ reliably for the two age groups.

Two separate 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on the intrusion rate were conducted, with the first ANOVA including only the study-only and one-guess conditions, and the second ANOVA including only the one-guess and three-guesses conditions. Critically, both ANOVAs again revealed a significant main effect of TYPE OF PRACTICE, all ps ≤ .004, reflecting that the overall intrusion rate was reduced in the one-guess condition, relative to the study-only condition, and was lower in the three guesses-condition, relative to the one-guess condition. Both ANOVAs showed no interaction between factors, all ps ≥ .13.

In Experiment 1, on average, about 36% of the intrusions made in the one-guess and three-guess conditions were incorrect guesses from the acquisition phase for both young and older adults. A 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on these "pretest" intrusions revealed no significant main effects and also no interaction between factors, all ps ≥ .09, which may be unsurprising given the relatively low frequency of this type of intrusion.

Discussion

The findings of Experiment 1 replicate previous research by demonstrating that the size of the pretesting effect can increase in young adults after three initial guesses, compared to a single initial guess (Kliegl et al., 2024). More important, the findings suggest that

Table 1
Mean Intrusion Rate Together With Standard Errors of the Mean (in Brackets) on the Final Test as a Function of Age Group and Type of Practice for Experiments 1 and 2

	Young adult			Older adult		
Experiment	Study	One guess	Three guesses	Study	One guess	Three guesses
Experiment 1						
Intrusion rate in %	26.00 (2.50)	19.50 (2.28)	18.00 (2.28)	34.25 (3.30)	30.42 (2.77)	26.42 (2.77)
Experiment 2						
Intrusion rate in %	19.44 (1.38)	15.12 (1.88)	12.03 (1.73)	27.16 (1.92)	20.37 (2.56)	22.84 (2.26)

older adults exhibit the same pattern of results. Specifically, older adults already showed a reliable pretesting effect of approximately 8% after only a single guess, which increased to approximately 14% after three guesses. The results also showed that, for both age groups, the number of intrusions produced on the final test decreased from study-only to one-guess trials and even from one-guess trials to three-guesses trials. The finding on intrusions is quite striking because older adults' episodic memory has been found to be particularly vulnerable to interference effects (e.g., Kliegl et al., 2019), which is why producing multiple erroneous guesses during pretesting rather could have been expected to lead to an increased number of intrusions. The observed effects of pretesting on both correct recall and intrusions may indicate that, similar to young adults, these guesses facilitate, rather than impair, access to the target information on the final test (see also General Discussion).

Experiment 2

The first goal of Experiment 2 was to determine whether the benefits of repeated guessing attempts during pretesting observed in Experiment 1 still arise for older adults when using study material with greater practical relevance (i.e., prose passages). To this end, both young and older participants studied three prose passages. One of the passages was presented in the study-only format, in which the sentences were displayed sequentially on a computer screen and could be learned immediately. The other two passages were employed for the one-guess and three-guesses formats. In these two conditions, sentences were initially presented in a fill-in-the-blank format and subjects were instructed to generate either one guess or three unique guesses about the missing key word before the complete sentence was revealed. Using such prose passages, one prior study has shown that, for young adults, three initial guessing attempts can induce a larger pretesting effect than a single initial guessing attempt (Kliegl et al., 2024). Since Experiment 1 demonstrated parallel increases in the size of the pretesting effect with repeated guessing attempts for weakly related word pairs as study material, a priori, one might assume that such parallel increases also arise for the two age groups when prose passages are used as study material.

The second goal was to investigate for the two age groups whether guessing during pretesting can also facilitate later retention of information that appeared in the passage but was initially not pretested. To this end, the fill-in-the-blank sentences in the one-guess and three-guesses formats alternated with complete sentences (i.e., the untested information). On the final test, participants' memory of both pretested and not pretested information was assessed. While prior work suggests that, oftentimes, pretesting fails to induce any transfer to untested material even in young adults (e.g., James & Storm, 2019; Kliegl et al., 2024; Richland et al., 2009), it has been demonstrated that transfer effects can arise when participants are explicitly instructed to study not only the pretested material but also untested information (e.g., Carpenter & Toftness, 2017). Therefore, we assumed that emphasizing the importance of studying all sentences for the later test might promote also final test performance of untested information for the two age groups.

Method

Participants

Fifty-four young adults with a mean age of 21.8 years (48 female, six male, zero diverse) and 54 older adults with a mean age of 66.6 years (36 female, 18 male, zero diverse) were recruited to take part in Experiment 2. For reasons of counterbalancing, six fewer participants were tested than in Experiment 1. We recruited young adults from Regensburg University and older adults from the community. All data were collected between September 2023 and February 2024. All participants were Caucasian and gave their spoken informed consent. Young participants were compensated either financially or earned some course credit for their participation. Older participants all received a compensatory amount of money for their participation. Compared to young adults, older adults performed significantly worse in the Mini Mental State Examination (young adults: 29.9; older adults: 28.9), $t(106) = 5.62$, $p < .001$, Cohen's $d = 1.08$, and the digit span task (young adults: 16.6; older adults: 15.5), $t(106) = 2.13$, $p = .035$, Cohen's $d = 0.41$, but significantly better in the MWT-B (young adults: 26.9; older adults: 32.8), $t(118) = 9.80$, $p < .001$, Cohen's $d = 1.89$,

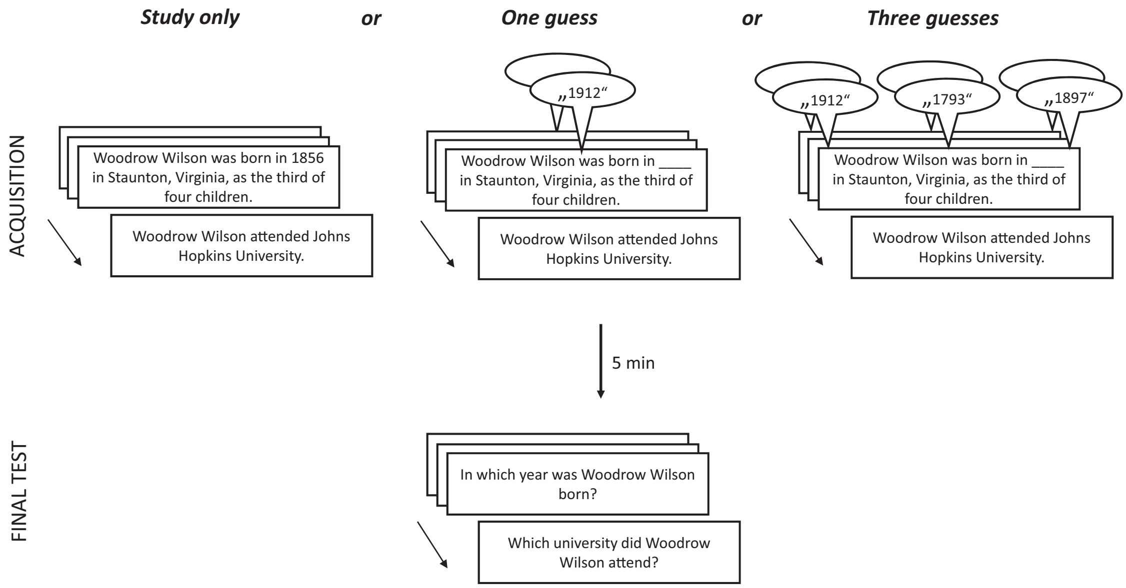
Material

All participants studied three short German text passages which had previously been used in Kliegl et al. (2024, Experiment 4) and dealt with the former U.S. President Woodrow Wilson (233 words), the country of Brazil (266 words), and Tapirs (252 words). The texts were created on the basis of information taken from Wikipedia (https://en.wikipedia.org/wiki/Woodrow/Wilson; https://en.wikipedia.org/wiki/Brazil; https://en.wikipedia.org/wiki/Tapir). For each of the three texts, understanding of their later sections did not depend on integration with their earlier sections (e.g., Kriechbaum & Bäuml, 2024; Wissman et al., 2011). The three texts consisted of 12 sentences and were employed equally often in each of the three practice conditions. For each sentence, a short-answer question was created for the final test which was based on information that appeared in the sentence (e.g., *When did Woodrow Wilson's presidency begin?*). We divided these 12 questions into two subgroups (Subgroups A and B), each of which contained six questions. For half of the young and older adults, the responses that were required for the questions of Subgroup A referred to the key word missing in the fill-in-the-blank sentences, while for the other half of young and older adults, the responses that were required for the questions of Subgroup B referred to the key word missing in these fill-in-the-blank sentences. On the final test, participants had to respond to all 12 questions.

Design and Procedure

The experiment again used a mixed design with the between-subjects factor AGE GROUP (young adults vs. older adults) and the within-subjects factor TYPE OF PRACTICE (study only vs. one guess vs. three guesses). In each of the three practice conditions, all participants first studied one of the three texts and, after completing a 5 min distractor activity, engaged in a final test on that passage (see Figure 2). During the acquisition phase of the study-only condition, the 12 sentences of a passage were shown one by one. Each sentence was presented for 25 s overall. During the acquisition phase of the one-guess condition, the 12 sentences of a passage were again

Figure 2
Procedure of Experiment 2



Note. While in the study-only condition, the 12 sentences of a passage were shown sequentially, in the one-guess and three-guesses conditions, the 12 sentences were also shown sequentially, but for six of the sentences, one key word was missing and participants had to make either one guess or three unique guesses before the missing word was revealed. Complete and fill-in-the-blank sentences were always presented alternatingly in the one-guess and three-guesses conditions. After this acquisition phase and a 5-min interval filled with unrelated distractor activities, participants engaged in a final test on all 12 sentences.

presented one by one at a rate of 25 s per sentence. However, only six of the sentences appeared intact from the beginning. For the other six sentences, a key word was missing (e.g., *Woodrow Wilson was born in ____ [year] in Staunton, Virginia, as the third of four children*.) and participants were informed 10 s after the onset of the sentence via a tone signal to make a guessing attempt. Subjects were given 5 s to attempt their guess before the key word was shown for 10 s. The acquisition phase of the three-guesses condition resembled the acquisition phase of the one-guess condition with the only exception being that participants were required to produce three unique guesses before the key word was shown. Participants had 5 s for each guess. A critical feature of the one-guess and three guesses conditions was that fill-in-the-blank sentences and complete sentences were always shown in an alternating manner. After the acquisition phase, participants engaged in a backward counting task for 1 min and in the standard progressive matrices task for another 4 min before the final test on the immediately preceding text was conducted. This test included the 12 questions on the text, with each question referring to one particular sentence of the text (e.g., *In which year was Woodrow Wilson born?*). Participants were required to answer each question within a time frame of 8 s. After this test, participants played Tetris (https://www.geo.de/geolino/spiele/13349-rtkl-onlinespiel-tetris) for 3 min prior to the presentation of the next text passage.

Results

Initial Pretest

A 2 × 2 ANOVA of the between-subjects factor AGE GROUP (young vs. older adults) and the within-subjects factor TYPE OF PRACTICE (one guess vs. three guesses) on percentage of correct guesses during the initial pretest showed a significant main effect of TYPE OF PRACTICE, $F(1, 106) = 8.09$, $p = .005$, $\eta_p^2 = .07$, reflecting that more correct guesses were made for item pairs for which participants were allowed to make three guessing attempts, relative to one guessing attempt (8.2% vs. 4.5%). There was no significant main effect of AGE GROUP and no significant interaction between factors, all Fs < 1.

Final Test

Correct Recall—Tested Information. Figure 1B shows the percentage of initially pretested questions for which correct answers were given on the final test as a function of the between-subjects factor AGE GROUP (young adults vs. older adults) and the within-subjects factor TYPE OF PRACTICE (study only vs. one guess vs. vs. three guesses). A 2 × 3 mixed ANOVA of the two factors on correct-recall performance revealed significant main effects of AGE GROUP, $F(1, 106) = 60.31$, $p < .001$, $\eta_p^2 = .36$, and TYPE OF PRACTICE, $F(2, 212) = 21.41$, $p < .001$, $\eta_p^2 = .17$, reflecting that overall recall performance of older adults was lower than of young adults (55.9% vs. 76.7%) and that overall recall performance was lower for the study-only condition than the one-guess and three-guesses conditions (58.9% vs. 70.1% vs. 69.8%). Critically, there was a significant interaction between the two factors, $F(2, 212) = 3.13$, $p = .046$, $\eta_p^2 = .03$, suggesting that the effects of TYPE OF PRACTICE varied with age group.

Two separate 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on correct-recall performance were conducted, with the first ANOVA including only the study-only and one-guess conditions, and the second ANOVA including only the one-guess and three-guesses conditions. The first ANOVA showed a significant main effect of TYPE OF PRACTICE, $F(1, 106) = 31.33$, $p < .001$, $\eta_p^2 = .23$, reflecting an overall increase in correct answers from the study-only to the one-guess condition. Furthermore, the first ANOVA showed no significant interaction between the two factors, $F(1, 106) = 1.79$, $p = .18$, $\eta_p^2 = .02$, reflecting that both age groups showed a similar increase in correct answers from the study-only to the one-guess condition. Critically, the second ANOVA showed no reliable main effect of TYPE OF PRACTICE, $F(1, 106) < 1$, $\eta_p^2 = .002$, reflecting no reliable difference in overall correct answers between the one-guess and the three-guesses conditions. In addition, this ANOVA did show a significant interaction between factors, $F(1, 106) = 7.96$, $p = .006$, $\eta_p^2 = .07$. Consistently, planned comparisons using the Wilcoxon signed-rank test revealed that young adults' correct recall increased reliably from the one-guess to the three-guesses condition (77.7% vs. 83.6%), $Z(53) = 2.01$, $p = .04$, whereas older adults' correct recall did not differ reliably between the one-guess and the three-guesses conditions (62.6% vs. 56.0%), $Z(53) = 1.73$, $p = .08$.

Correct Recall—Untested Information. Figure 1B shows the percentage of initially untested questions for which correct answers were given on the final test. A 2 × 3 mixed ANOVA on correct-recall performance revealed significant main effects of AGE GROUP, $F(1, 106) = 50.44$, $p < .001$, $\eta_p^2 = .32$, and TYPE OF PRACTICE, $F(2, 212) = 5.47$, $p = .004$, $\eta_p^2 = .05$, reflecting that overall recall performance of older adults was lower than of young adults (45.7% vs. 65.7%) and that overall recall performance was lower for the one-guess condition than the study-only and three-guesses conditions (58.9% vs. 51.7% vs. 56.4%).

Two separate 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on correct-recall performance were conducted, with the first ANOVA including only the study-only and one-guess conditions, and the second ANOVA including only the one-guess and three-guesses conditions. Both ANOVAs revealed significant main effects of TYPE OF PRACTICE, all $ps \leq .036$, reflecting overall lower recall performance in the one-guess than study-only condition for the first ANOVA and overall higher recall performance in the three-guesses than one-guess condition for the second ANOVA. Neither of the two ANOVAs showed a significant interaction between factors, all $ps \geq .45$. A third 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on correct-recall performance which included only the study-only and three-guesses condition showed neither a main effect of TYPE OF PRACTICE, $F(1, 106) = 1.64$, $p = .20$, $\eta_p^2 = .02$, nor an interaction between factors, $F(1, 106) < 1$, $\eta_p^2 = .001$.

Intrusions. Since the analysis of correct-recall performance did not provide any evidence of transfer effects, we report analyses of intrusion errors only for the tested information. Table 1 shows intrusion rates as a function of AGE GROUP and TYPE OF PRACTICE. A 2 × 3 mixed ANOVA of the two factors on intrusion rate revealed significant main effects of AGE GROUP, $F(1, 106) = 16.32$, $p < .001$, $\eta_p^2 = .13$, and TYPE OF PRACTICE, $F(2, 212) = 9.74$, $p < .001$, $\eta_p^2 = .08$, reflecting that overall intrusion rate of older adults was higher than of young adults (23.5% vs. 15.5%) and that overall intrusion rate was higher for the study-only condition than the one-guess and three-guesses conditions (23.3% vs. 17.7% vs. 17.4%). There was no significant interaction between the two factors, $F(2, 212) < 1$, $\eta_p^2 = .007$.

Two separate 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on intrusion rate were conducted, with the first ANOVA including only the study-only and one-guess conditions, and the second ANOVA including only the one-guess and three-guesses conditions. Critically, only the first ANOVA showed a significant main effect of TYPE OF PRACTICE, $F(1, 106) = 12.58$, $p < .001$, $\eta_p^2 = .11$, whereas the second ANOVA did not, $F(1, 106) < 1$, $\eta_p^2 = .001$, reflecting an overall decrease in intrusions from the study-only to the one-guess condition but not from the one-guess to the three-guesses condition. Neither of the two ANOVAs showed a significant interaction between factors, all $ps \geq .29$.

In Experiment 2, on average, about 9% (young adults) and 8% (older adults) of the intrusions made in the one-guess and three-guess conditions were incorrect pretest guesses. The relatively low percentage of pretest intrusions in Experiment 2 compared to Experiment 1 may indicate that the type of study material influences the probability of "pretest" intrusions, with semantically richer prose passages (Experiment 2) potentially making participants less susceptible to this type of commission error than weakly related word pairs (Experiment 1). Like in Experiment 1, a 2 × 2 mixed ANOVAs of AGE GROUP and TYPE OF PRACTICE on "pretest" intrusions did not reveal any reliable main effects or an interaction, all $ps \geq .65$.

Discussion

The results of Experiment 2 showed a pretesting effect for both young and older adults when a single guess had been made initially, thereby replicating the findings of the current Experiment 1. However, only young adults were found to benefit from three initial guessing attempts, whereas older adults even showed a numerical, but not statistically significant, decrease in final test recall after three guesses, relative to a single guess. Critically, older adults' final test recall of pretested information following three initial guesses was still greater than following study only, suggesting that older adults did benefit, at least to some degree, from multiple initial guesses. The results on intrusion rates suggest that pretesting again reduced intrusions on the final test for both age groups, but that the added reduction in intrusions was not statistically reliable when three guesses, instead of one guess, had to be made.

Furthermore, the current results suggest that for both age groups, pretesting was unable to enhance subsequent memory performance of information that was not pretested. This held both in response to a single initial guess and three initial guesses, which aligns with the results of a recent study showing also no benefits for young adults of either one or three initial guesses on later recall of untested information (Kliegl et al., 2024). The current findings suggest that even emphasizing to participants that all details contained within the prose passages would be relevant for the final test does not necessarily trigger a transfer effect of pretesting.

General Discussion

The present study demonstrates that both young and older adults can benefit from pretesting even when multiple initial guesses are made during acquisition. For young adults, the results of Experiments 1 and 2 showed that, with both weakly associated word pairs and prose passages as study material, repeated initial guessing attempts can increase the size of the pretesting effect on the final test—thus replicating prior work (Kliegl et al., 2024). However, final test recall in Experiment 2 did not provide evidence for a pretest-induced transfer effect to previously untested but studied information (for a more detailed discussion, see below). As for young adults, Experiment 1 showed a reliable pretesting effect for older adults after one initial guessing attempt which increased in magnitude when three initial guesses had to be made. Experiment 2 also showed mostly parallel effects of pretesting between young and older adults, with reliable pretesting effects following both a single guess and three guesses, as well as a lack of any transfer effects to previously untested but studied information. Unlike for young adults, however, Experiment 2 found no additional benefit of repeated guessing attempts during pretesting for older adults.

The Effects of Pretesting on Tested Information

The current results on the effects of pretesting on tested material are well in line with the elaboration account for young adults. According to this explanation of the pretesting effect, initial guessing attempts should induce a more elaborate activation of memory representations related to the cue item than study only, thus facilitating access to the target item on the final test. Consequently, repeated guessing attempts during pretesting would be expected to create even more memory traces than a single guessing attempt through which retrieval of the target item is enabled on the final test, thereby increasing the magnitude of the pretesting effect. For young adults, both Experiments 1 and 2 found that the magnitude of the pretesting effect increased with the number of initial guessing attempts, which is exactly what the account predicts. For older adults, only Experiment 1—but not Experiment 2—demonstrated that the magnitude of the pretesting effect increased with number of guesses, which appears more difficult to align with the elaboration view (see below).

While most of the present results for tested material are broadly consistent with the elaboration account of the pretesting effect, most of these results also align with another prominent explanation of the effect—the episodic context account (Metcalfe & Huelser, 2020). This account proposes that when we store information, we encode features of the temporal context that is present when we encounter that information (Estes, 1955; Mensink & Raaijmakers, 1988). These features may relate to various aspects of an event, such as the external environment or a person's mental state. While temporal context is assumed to change over time, there is evidence that certain activities, such as memory retrieval, can accelerate contextual shifts (e.g., Jang & Huber, 2008; Jonker et al., 2013). Since pretesting should involve higher levels of memory retrieval than studying alone, pretested information can get linked to a richer context that includes an integrated memory representation of the pretest context and the study context, whereas study-only information should be linked with merely the study context. Repeated guessing attempts during pretesting may afford an additional benefit because each guessing triggers a slightly new context and therefore a wider variety of contextual elements that become associated with the pretested information, thereby promoting recall of the pretested information on the final test. While the account resembles the elaboration account by assuming that repeated guesses induce a richer memory representation than a single guess, critically, this enrichment is thought to be episodic—and not semantic—in nature. The account thus can also explain the beneficial effects of repeated guessing attempts on later recall. However, as with the elaboration account, it is also not immediately obvious how it should explain why older adults only showed an increased pretesting effect when the number of guesses was increased in Experiment 1, but not in Experiment 2.

The discrepancy in results between weakly related word pairs (Experiment 1) and more factual text-based material (Experiment 2) may remind us of earlier work on possible differences in pretesting effects between word pairs and text-based material, in which immediate correct-answer feedback was required to induce the effect with weakly associated word pairs, whereas delayed feedback produced the effect with trivia questions or prose passages (e.g., Kornell, 2014; Richland et al., 2009). This discrepancy has been attributed, for instance, to the possibility that pretest cues (i.e., questions) which refer to text-based material—such as *What year was Woodrow Wilson born?*—may generally be more memorable after a delay than pretest cues which refer to an associated cue word (e.g., star—? Kornell, 2014). Regarding the present discrepancy in results obtained for word pairs and prose passages, there may be a critical role for older adults' well-known susceptibility to intralist interference (e.g., Radvansky et al., 2005)—a type of interference arising from the competition between elements of a single study list or segment (e.g., Raaijmakers & Shiffrin, 1981; Rundus, 1973).

Indeed, while the weakly related word pairs used in Experiment 1 have been shown in earlier work to induce relatively low levels of interference between the pairs of a list (Tulving & Arbuckle, 1963), the fact that the sections of the text passages used in Experiment 2 were difficult to integrate but all revolved around the same topic (e.g., *Woodrow Wilson*) without being easily integrated makes it possible that the text passages had considerable potential to induce intralist interference. This interference potential may have been particularly high on the final test when participants were asked to recall information for which three guesses had been made during pretesting. In these cases, participants not only had their own guesses as a possible source of interference but also facts from the other sections of the text. For instance, the final test question *What year was Woodrow Wilson born?* (correct answer: *1856*) may not only have elicited earlier guesses (e.g., *1920*, *1834*, *1899*) but also, for example, the year in which Wilson enrolled at his university (i.e., *1883*). Due to their increased susceptibility to intralist interference, older adults in particular may have had problems at accessing the correct response on the final test, which could explain why repeated guessing attempts induced no benefits in older adults in Experiment 2. It is still possible that repeated guessing attempts could benefit memory in older adults when using texts whose subsections can be integrated more easily than in the present study (e.g., Kriechbaum & Bäuml, 2024; Wissman et al., 2011), since integration has been shown to reduce interference (e.g., Wahlheim & Jacoby, 2013).

The (Lack of) Effects of Pretesting on Untested Information

If pretesting indeed enhances elaborative encoding, and if this type of processing spills over to untested information, then recall performance of untested information should also improve after three guesses compared to one guess. For young and older adults, the results of Experiment 2 did reveal such a pattern, which aligns with the elaboration view. However, these considerations would also lead to the expectation that recall of untested information is higher after a single guessing trial—and even more so after three trials—than in the study-only condition, which was not the case in Experiment 2. The present results thus did not provide any evidence for pretest-induced transfer effects to untested information for either young or older adults. We had expected that by i) emphasizing to participants prior to the start of the experiment that all study material would be relevant on the final test and ii) presenting untested sentences in close temporal proximity to sentences which received a pretest during the acquisition phase might promote such transfer. While the finding appears inconsistent with Carpenter and Toftness (2017) demonstrating transfer effects when all study material was deemed important for the later test, it aligns with the observations from several studies which also failed to provide any evidence for pretest-induced transfer effects to untested information (e.g., James & Storm, 2019; Kliegl et al., 2024; Richland et al., 2009).

Similar to the elaboration account, the episodic context account cannot easily accommodate for the detrimental effect of pretesting on the recall of untested information after a single guessing attempt. Unlike the elaboration account, however, the episodic context account also cannot explain why three guessing attempts improve the recall of untested information relative to a single guessing attempt. Indeed, the episodic context account assumes that pretest-induced enrichments in mental context should only affect tested information, irrespective of the number of guessing trials. In its current form, the account therefore appears less capable of explaining the current pattern of results for both tested and untested information than the elaboration account.

The recently proposed attentional-window hypothesis (Sana & Carpenter, 2023) might explain why we failed to obtain a transfer effect in Experiment 2. The account applies particularly to pretest situations in which texts or videos are used as study material and participants are asked to answer a number of pretest questions before the material is exposed. The assumption is that such pretesting opens an attentional window in which participants focus on finding the correct responses to the pretest questions. Critically, the window is assumed to close once the correct answer has been found. In support of their hypothesis, Sana and Carpenter showed that final test recall of untested information gets boosted when the information appears "in the attentional window," that is, prior to the tested information, but not when the untested information appears after the tested information. As the untested information in Experiment 2 was always presented after the tested information, it seems plausible on the basis of this explanation that no transfer effect was observed. Critically, the attentional-window and elaboration accounts are not mutually exclusive because, for instance, a pretest-induced attentional window might trigger a more elaborative encoding of untested information that is encountered prior to the pretested information. It should be a high priority of future research to investigate such a possibility.

A Practical Recommendation

From a more practical perspective, the current findings, together with prior work (Cyr & Anderson, 2012, 2015) suggest that pretesting may serve as an effective learning tool not only in young adults but also in older adults. The results of the present two experiments in particular indicate that, for both age groups, an initial learning session which includes at least a single guessing attempt can promote subsequent recall performance more effectively than a learning session which includes mere study of the information. Another, more preliminary, conclusion from the present findings might be that repeated initial guessing attempts may further boost older adults' memory, as long as a relatively simple learning task is used, in which a cue information needs to be encoded together with a target information without distraction from other to-be-learned material. This might include such tasks as vocabulary learning or the learning of face-name pairs. More complex material like prose passages may limit the pretesting effect—especially when its sections are difficult to integrate—and not lead to a boost in performance with multiple guessing attempts.

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

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., & 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

Clare, L., & Jones, R. S. (2008). Errorless learning in the rehabilitation of memory impairment: A critical review. *Neuropsychology Review*, *18*, 1–23. https://doi.org/10.1007/s11065-008-9051-4

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

Cyr, A. A., & Anderson, N. D. (2015). Mistakes as stepping stones: Effects of errors on episodic memory among younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 841–850. https://doi.org/10.1037/xlm0000073

Duker, P. C. (1981). Prevention of incorrect responding for establishing instruction following behaviours. *Journal of Mental Deficiency Research*, *25*(1), 25–32. https://doi.org/10.1111/j.1365-2788.1981.tb00090.x

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

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

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., & 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

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 112–127. https://doi.org/10.1037/0278-7393.34.1.112

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

Kang, S. H., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*(1), 48–59. https://doi.org/10.1037/a0021977

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

Kay, M., Elkin, L. A., Higgins, J. J., & Wobbrock, J. O. (2021). *ARTool: Aligned rank transform. [R] package with documentation and two vignettes*.

Keel, M. C., & Gast, D. L. (1992). Small-group instruction for students with learning disabilities: Observational and incidental learning. *Exceptional Children*, *58*(4), 357–368. https://doi.org/10.1177/001440299205800408

Kliegl, O., Bartl, J., & Bäuml, K. H. T. (2024). Repeated guessing attempts during acquisition can promote subsequent recall performance. *Journal of Experimental Psychology: Applied*, *30*(2), 282–292. https://doi.org/10.1037/xap0000493

Kliegl, O., Carls, T., & Bäuml, K.-H. T. (2019). How delay influences search processes at test. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(12), 2174–2187. https://doi.org/10.1037/xlm0000701

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

Kriechbaum, V. M., & Bäuml, K.-H. T. (2024). Retrieval practice can promote new learning with both related and unrelated prose materials. *Journal of Applied Research in Memory and Cognition*, *13*(3), 319–328. https://doi.org/10.1037/mac0000162

Lehrl, S. (2005). *MWT-B Mehrfachwahl. Wortschatz-Intelligenztest* [Multiple-Choice Vocabulary Intelligence Test (MWT-B)]. Spitta-Verlag.

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

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*, 225–229. https://doi.org/10.3758/BF03194056

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/

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review*, *35*(4), Article 97. https://doi.org/10.1007/s10648-023-09814-5

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

Radvansky, G. A., Zacks, R. T., & Hasher, L. (2005). Age and inhibition: The retrieval of situation models. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), 276–278. https://doi.org/10.1093/geronb/60.5.P276

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. Psychology Press.

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

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

Rundus, D. (1973). Negative effects of using list items as retrieval cues. *Journal of Verbal Learning and Verbal Behavior*, *12*, 43–50. https://doi.org/10.1016/S0022-5371(73)80059-3

Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review*, *30*(5), 1908–1916. https://doi.org/10.3758/s13423-023-02274-6

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

St. Hilaire, K. J., Chan, J. C., & Ahn, D. (2024). Guessing as a learning intervention: A meta-analytic review of the prequestion effect. *Psychonomic Bulletin & Review*, *31*(2), 411–441. https://doi.org/10.3758/s13423-023-02353-8

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*(5), 321–334. https://doi.org/10.1016/S0022-5371(63)80012-2

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition*, *41*, 1–15. https://doi.org/10.3758/s13421-012-0246-9

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

Wilson, B. A., Baddeley, A., Evans, J. J., & Shiel, A. (1994). Errorless learning in the rehabilitation of memory impaired people. *Neuropsychological Rehabilitation*, *4*(3), 307–326. https://doi.org/10.1080/09602019408401463

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, *18*, 1140–1147. https://doi.org/10.3758/s13423-011-0140-7

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the ACM conference on human factors in computing systems (CHI '11)* (pp. 143–146). Association for Computing Machinery.

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

Received December 19, 2024
Revision received June 26, 2025
Accepted June 27, 2025 ■